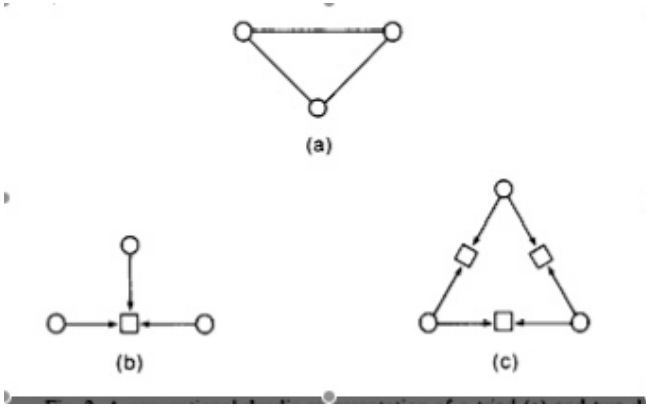
Unit 7: Affiliations Data and Bibliographic Research

ICPSR
University of Michigan, Ann Arbor
Summer 2015
Instructor: Ann McCranie

"Affiliations"



Relational in three ways
(Wasserman and Faust, pg 295)

1. Show how actors and events are related to one another
2. Events create ties among actors
3. Actors create ties among events

Note: You are no longer considering pairs of actors, but instead are considering subsets of actors.

Classic Study: Southern Ladies

Davis, Gardner, Gardner 1941. *Deep South*: a Social Anthropological Study of Caste and Class

See also: Freeman, Linton. 2003 Finding Social Groups: A Meta-Analysis of the Southern Women Data)

		1 E	2 E	3 E	4 E	5 E	6 E	7 E	8 E	9 E	10 E	11 E	12 E	13 E	14 E
1	EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
2	LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
3	THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
4	BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
5	CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
6	FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
7	ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
8	PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
9	RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
10	VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
11	MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
12	KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
13	SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
14	NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
15	HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
16	DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
17	OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
18	FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

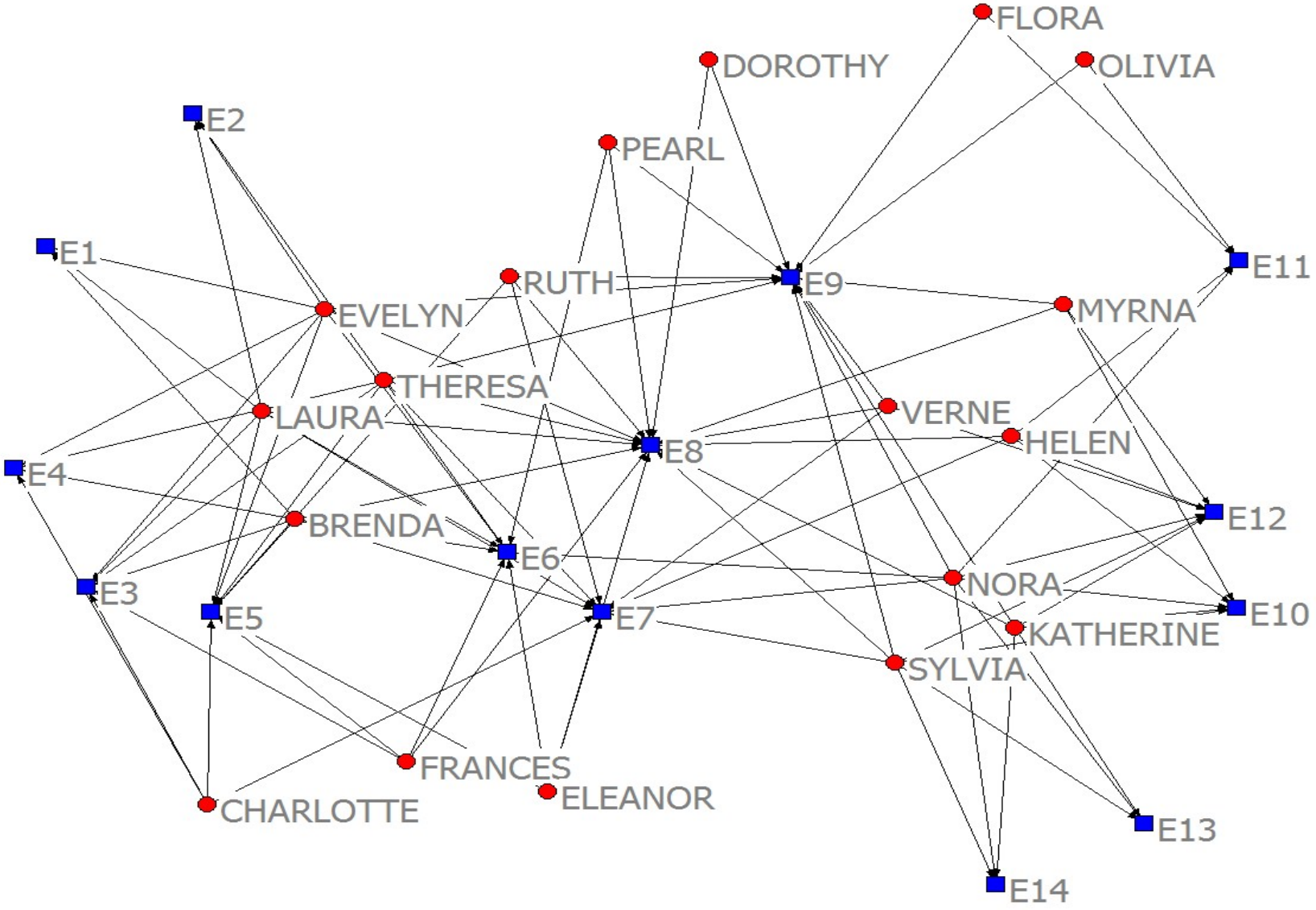
Two-Mode Data Issues

Think carefully about the assumptions behind the collection of actors and events.

Does belonging to the same organization count as an affiliation? What about belonging to the same type of organization? How about sharing the same attitude?

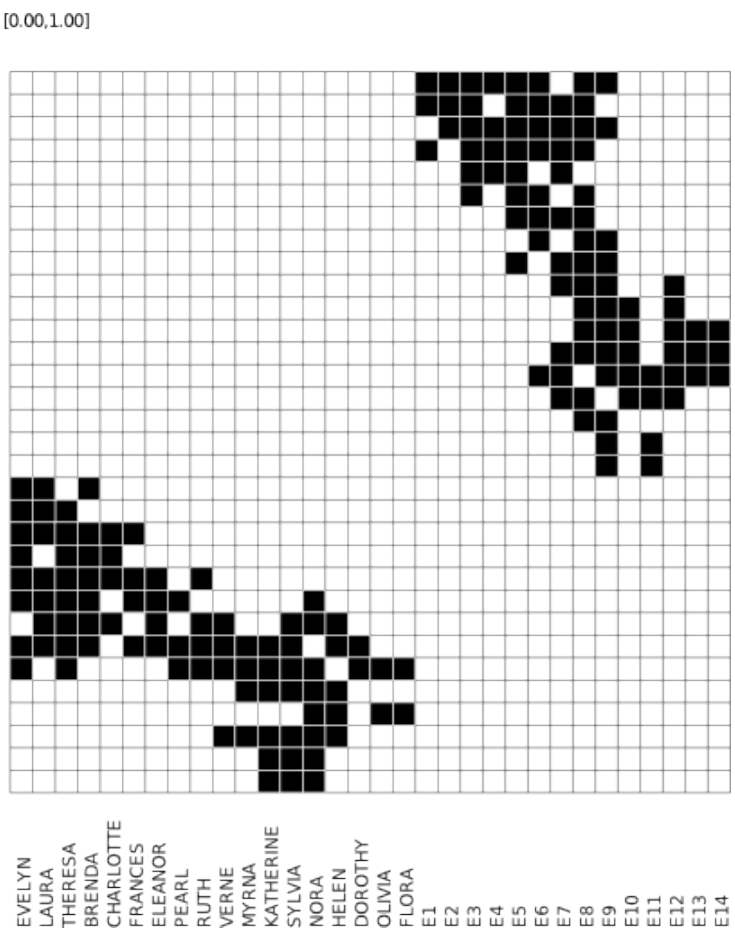
Also, deciding on the population boundaries *of each* is important.

Duality of Actors and Events



Two-Mode Centrality: Bipartite Matrix

		1 E	2 L	3 T	4 B	5 C	6 F	7 E	8 P	9 R	10 V	11 M	12 K	13 S	14 N	15 H	16 D	17 O	18 F	19 E	20 E	21 E	22 E	23 E	24 E	25 E	26 E	27 E	28 E	29 E	30 E	31 E	32 E
1	EVELYN																			1	1	1	1	1	1		1	1						
2	LAURA																			1	1	1		1	1	1	1							
3	THERESA																				1	1	1	1	1	1	1	1						
4	BRENDA																			1		1	1	1	1	1	1							
5	CHARLOTTE																					1	1	1		1								
6	FRANCES																					1		1	1		1							
7	ELEANOR																							1	1	1	1							
8	PEARL																								1		1	1						
9	RUTH																							1		1	1							
10	VERNE																									1	1	1		1				
11	MYRNA																										1	1	1	1				
12	KATHERINE																										1	1	1	1	1	1	1	
13	SYLVIA																									1	1	1	1	1	1	1	1	
14	NORA																								1	1		1	1	1	1	1	1	
15	HELEN																									1	1		1	1	1			
16	DOROTHY																										1	1						
17	OLIVIA																										1		1					
18	FLORA																										1		1					
19	E1	1	1		1																													
20	E2	1	1	1																														
21	E3	1	1	1	1	1	1																											
22	E4	1		1	1	1																												
23	E5	1	1	1	1	1	1	1		1																								
24	E6	1	1	1	1		1	1	1						1																			
25	E7		1	1	1	1		1		1	1			1	1	1																		
26	E8	1	1	1	1		1	1	1	1	1	1	1	1		1	1																	
27	E9	1		1					1	1	1	1	1	1	1		1	1	1															
28	E1											1	1	1	1	1																		
29	E11														1	1		1	1															
30	E12										1	1	1	1	1	1																		
31	E13												1	1	1																			
32	E14												1	1	1																			



Two-Mode Data: Basic Measures

Properties you should know: Rates of participation (W&F 312-313) a size of events (W & F313-314)

UCINET now has some nice measures you can easily compute:

2-Mode Cohesion Measures for davis dataset.

	1 Density	2 Avg Dist	3 Radius	4 Diameter	5 Fragmenta	6 Transitiv	7 Norm Dist
Matrix 1	0.353	2.306	3.000	4.000	0.000	0.619	0.647

NOTE: If fragmentation is > 0, the graph is disconnected. All measures based on lengths of geodesics are computed within components.

Density is the number of ties divided by n*m, where these no. of rows and cols in matrix.
Avg Dist is the average geodesic path length in the bipartite graph, within components.
Radius is the smallest eccentricity in the bipartite graph, within components.
Diameter is the length of the longest geodesic in the bipartite graph, within components.
Transitivity is the no. of quadruples with 4 legs divided by no. with 3 or more legs, in bipartite graph.
Norm Dist is Avg Dist divided into minimum possible in bipartite graph of given node-set sizes.

Two-Mode Data: Network Density

For the “actors,” density is the average number of events attended by all pairs of actors.

$$D^A = \frac{\sum_{i=1}^{g}\sum_{j=1}^{g}\mathbf{X}_{ij}^{A}}{\frac{g(g-1)}{2}} \qquad (i < j)$$

You can do something similar for event density.

Two-Mode Data: Network Centrality

Actor Degree Centrality in an affiliation network is the total number of actors contacts that the actor (*i*) has through its attendance at all events.

Sum that actor's row in the coattendence matrix. (You can also do something similar for events.)

$$C_D^A(a_i) = \sum_{i=1}^{g} x_{ij}^A \qquad (i \neq j)$$

Centrality for Davis

2-Mode Centrality Measures for ROWS of davis

		1	2	3	4
		Degree	Closeness	Betweenne	Eigenvect
1	EVELYN	0.571	0.800	0.097	0.335
2	LAURA	0.500	0.727	0.051	0.309
3	THERESA	0.571	0.800	0.088	0.371
4	BRENDA	0.500	0.727	0.049	0.313
5	CHARLOTTE	0.286	0.600	0.011	0.168
6	FRANCES	0.286	0.667	0.011	0.209
7	ELEANOR	0.286	0.667	0.009	0.228
8	PEARL	0.214	0.667	0.007	0.180
9	RUTH	0.286	0.706	0.017	0.236
10	VERNE	0.286	0.706	0.016	0.218
11	MYRNA	0.286	0.686	0.016	0.187
12	KATHERINE	0.429	0.727	0.047	0.220
13	SYLVIA	0.500	0.774	0.072	0.277
14	NORA	0.571	0.800	0.113	0.264
15	HELEN	0.357	0.727	0.042	0.201
16	DOROTHY	0.143	0.649	0.002	0.131
17	OLIVIA	0.143	0.585	0.005	0.070
18	FLORA	0.143	0.585	0.005	0.070

2-Mode Centrality Measures for COLUMNS of davis

		1	2	3	4
		Degree	Closeness	Betweenne	Eigenvect
1	E1	0.167	0.524	0.002	0.142
2	E2	0.167	0.524	0.002	0.150
3	E3	0.333	0.564	0.018	0.253
4	E4	0.222	0.537	0.008	0.176
5	E5	0.444	0.595	0.038	0.322
6	E6	0.444	0.688	0.065	0.328
7	E7	0.556	0.733	0.130	0.384
8	E8	0.778	0.846	0.244	0.507
9	E9	0.667	0.786	0.226	0.379
10	E10	0.278	0.550	0.011	0.170
11	E11	0.222	0.537	0.020	0.090
12	E12	0.333	0.564	0.018	0.203
13	E13	0.167	0.524	0.002	0.113
14	E14	0.167	0.524	0.002	0.113

These are routines available in UCINET under “Network->2-Mode Networks and provide appropriate normalizations for the values. (If you just forced your data into a bipartite graph it would not be normalized correctly for the number of actors and events.)

Closeness and Betweeness

We can also get closeness and betweenness measures for two-mode networks, but we have to change the way they are normalized in order to reflect the fact that we have nodes that, by definition, can not be adjacent to one another.

Read Extending Centrality by Everett and Borgatti in Carrington for more, but think of this.

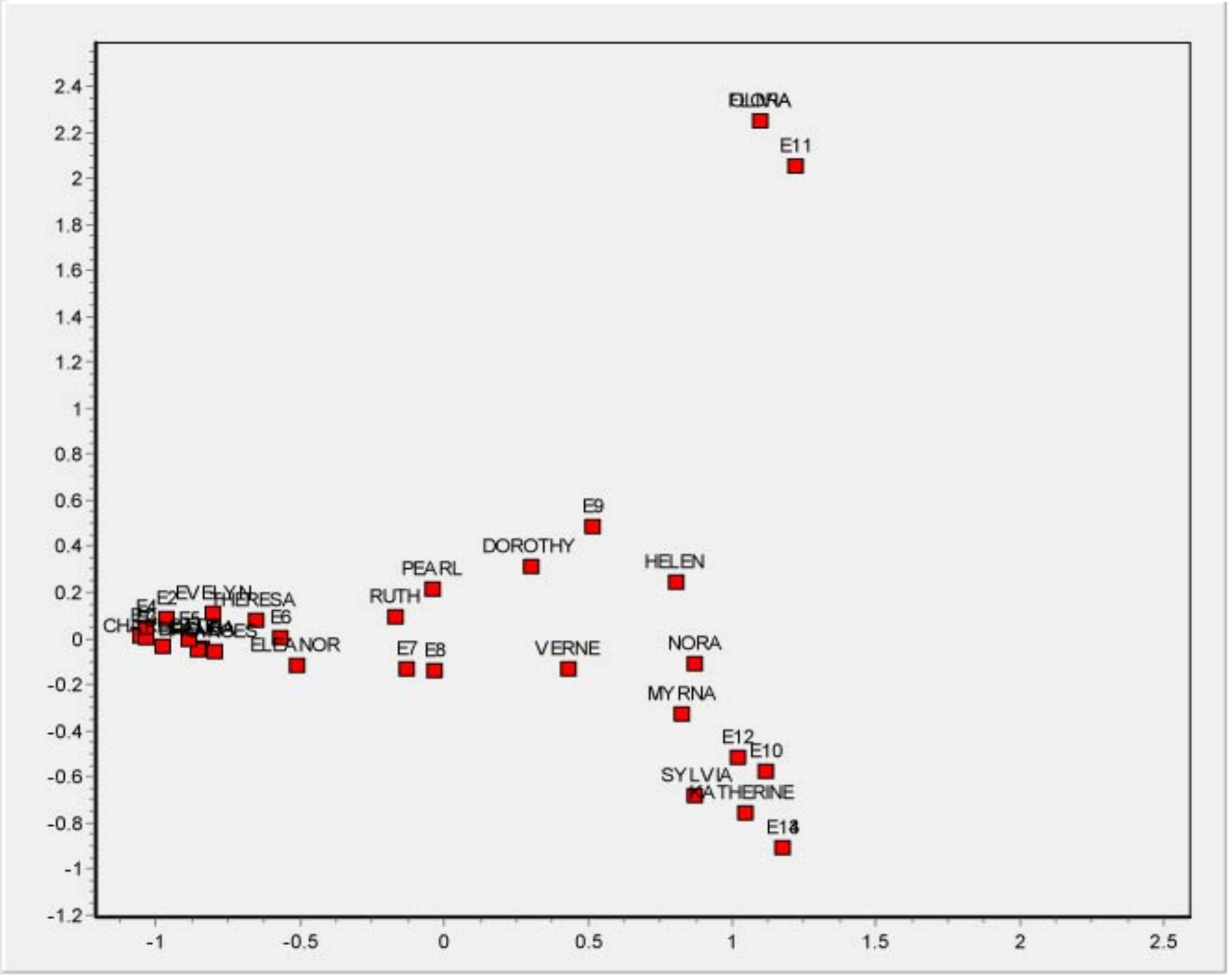
Actor *i* closeness centrality: $\frac{h + 2g - 2}{C_c(i)}$

Correspondence Analysis

(See Faust, Chapter 7 in Carrington, for excellent description)

Correspondence Analysis looks at the correlations between two sets of variables and is used to locate actors and events simultaneously: actors near events they attended and events near actors who attended them.

Correspondence analysis uses singular value decomposition of a normalized version of your *g x h* matrix


2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1
-0.5
0
0.5
1
1.5
2
2.5
E11
E9
DOROTHY
HELEN
PEARL
RUTH
E6
ELEANOR
E7
E8
VERNE
NORA
MYRNA
E12
E10
SYLVIA
THERINE

Factions in 2-mode data

Starting fitness: 0.000
Final fitness: 0.490
Correlation to ideal: 0.490

Blocked Adjacency Matrix

		1 E1	2 E2	3 E3	4 E4	5 E5	6 E6	7 E7	8 E8	9 E9	10 E10	11 E11	12 E12	13 E13	14 E14
1	EVELYN	1.000	1.000	1.000	1.000	1.000	1.000		1.000	1.000					
2	LAURA	1.000	1.000	1.000		1.000	1.000	1.000	1.000						
3	THERESA		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000					
4	BRENDA	1.000		1.000	1.000	1.000	1.000	1.000	1.000						
5	CHARLOTTE			1.000	1.000	1.000		1.000							
6	FRANCES			1.000		1.000	1.000		1.000						
7	ELEANOR					1.000	1.000	1.000	1.000						
8	PEARL						1.000		1.000	1.000					
9	RUTH					1.000		1.000	1.000	1.000					
10	VERNE							1.000	1.000	1.000			1.000		
11	MYRNA								1.000	1.000	1.000		1.000		
12	KATHERINE								1.000	1.000	1.000		1.000	1.000	1.000
13	SYLVIA							1.000	1.000	1.000	1.000		1.000	1.000	1.000
14	NORA						1.000	1.000		1.000	1.000	1.000	1.000	1.000	1.000
15	HELEN							1.000	1.000		1.000	1.000	1.000		
16	DOROTHY								1.000	1.000					
17	OLIVIA									1.000		1.000			
18	FLORA									1.000		1.000			

Density matrix

	1	2
1	0.625	0.074
2	0.153	0.537

Turning Two-Mode into One-Mode Symmetric Valued Data

Most common way to convert is called "cross-products." This method takes each entry of the row for actor A, and multiplies it times the same entry for actor B, and then sums the result.

- For binary data this ends up with a measure of concurrent attendance/membership.
- For valued data, you can use a "minimums" method: taking the smallest value of a pair of actor's ties to events and summing that with all other actors.

Other places to go…

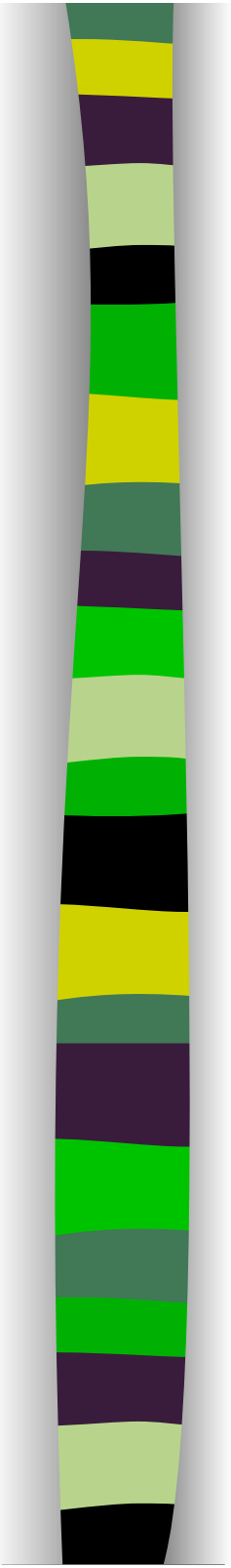
- Knoke and Yang: Social Network Analysis (2008)
- Faust: Using Correspondence Analysis for Joint Displays of Affiliation Networks (in Carrington, Scott and Wasserman)
- Roberts, J. M. Correspondence analysis of two-mode network data. Social Networks, 22:65-72, 2000.
- "Finding Social Groups: A Meta-analysis of the Southern Women Data" In Ronald Breiger, Kathleen Carley and Philippa Pattison (eds.) Dynamic Social Network Modeling and Analysis. Washington, D.C.:The National Academies Press, 2003.
- Borgatti, S. P. and M. G. Everett. Network analysis of 2-mode data. Social Networks, 19(3):243-269, 1997.

Recent extensions to 2-mode data

- Blockmodeling
  - P Doreian, V Batagelj..., Generalized blockmodeling of two-mode network data Social Networks, 200
- Large two-mode networks
  - Matthieu Latapy, Clemence Magnien, Nathalie Del Vecchio, Basic notions for the analysis of large two-mode networks, Social Networks, Volume 30, Issue 1, January 2008, Pages 31-48
- ERGM/p* models
  - P. Wang, K. Sharpe, G. Robins and P. Pattison, Exponential random graph models for affiliation networks, Social Networks 31 (1) (2009), pp. 12–25

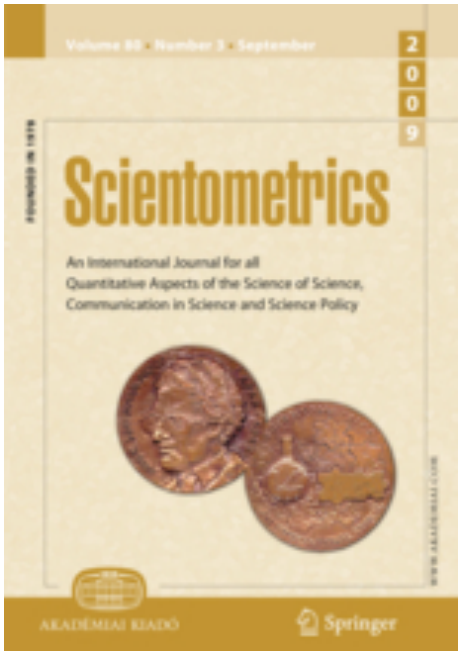
A demonstration and reminder

- Sci2
- There data available for you to download on the course webpage.



AN INTRODUCTION TO BIBLIOGRAPHIC NETWORK ANALYSIS

Scientometrics is an independent and diverse field


Scientometrics
An International Journal for all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy
2009
AKADÉMIAI KIADÓ
Springer

- Literally, measuring and analyzing science (all kinds).
- Today we're just in a tiny little corner of "bibliometric" analysis
  - Co-authorship
  - Citation
- But this field covers all sorts of ways of quantifying and evaluating scientific efforts.

Bibliographic Network Analysis

- Is often focused on citation or authorship patterns that can be found within fields
- Co-authorship is pretty clear – shared authorship of an article
- Citation analysis can be thought of differently. Two common ways:
  - “bibliometric coupling, where two citing articles are similar to the extent they cite the same literature, and co-citation analysis where cited articles are similar to the extent they are cited by the same citing articles” Hummon and Doriean Social Networks Volume 11, Issue 1, March 1989, Pages 39-63
- Citation networks are an interesting case, because at their most basic they are directed, acyclic data. (A paper cites a paper that was written earlier, that older paper can never cite anything published later.)

Bibliographic network analysis

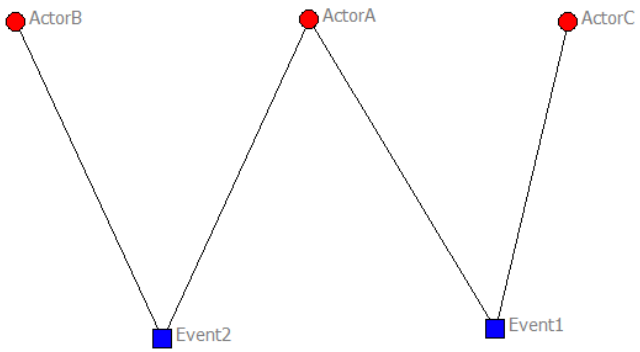
- Can detect subcommunities in research fields
- Can (and has been) be used to “map science”
- Can be used to identify prominent actors in a research field
- Can be used to identify conflict and consensus in science
- Can be used to trace the development of ideas

Why might you want to do this if you aren't into bibliometrics?

- Get a handle on an intellectual field
  - The people in that field
  - The academic work they produce and what is most cited
  - The clusters and divisions of the people and ideas in the field
- Maybe you just want to find the biggest stars and greatest hits – there are now a lot of tools to help you do that.

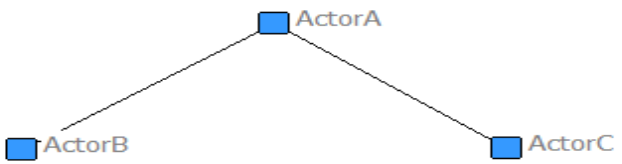
Studying Bibliographic Networks

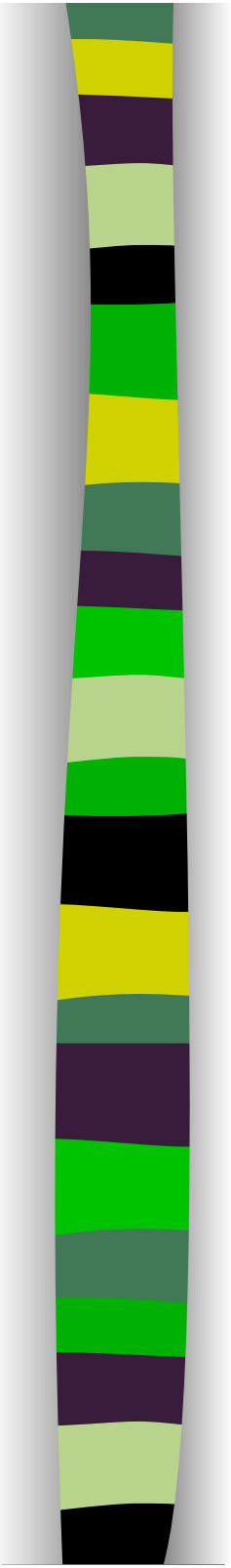
- These are special types of social networks
- Instead of two scholars being linked together by their relationship (such as being friends or naming each other as colleagues), two scholars are linked in a bibliographic network by the work they produce together.
- In a citation network, two papers are linked by the works they both cite or the works that both cite them.
- These are a type of “affiliation” network, technically a type of “two-mode” network.



Two Mode Network Actors & Events

Two Mode turned into One Mode Network (Just Actors)



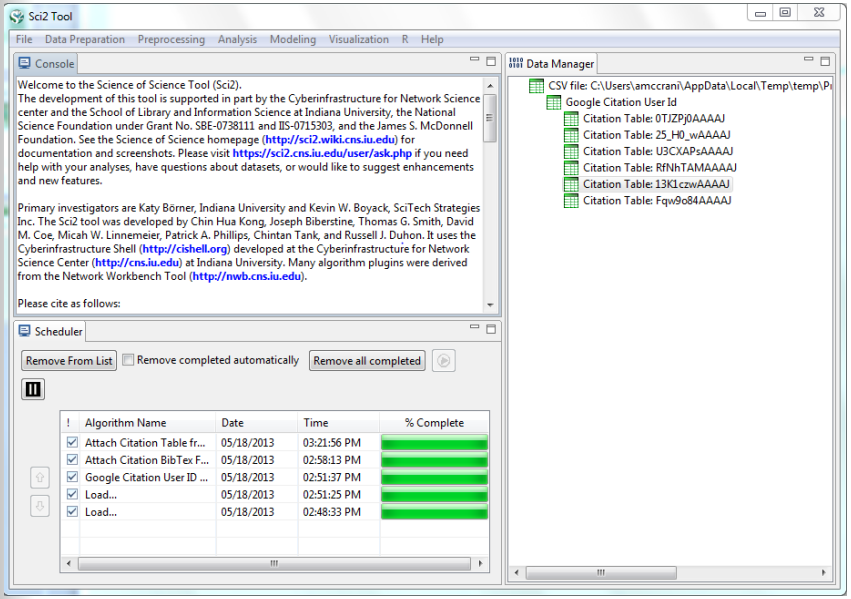


SCI² TOOL DEMONSTRATION

Download the tool and documentation:
https://sci2.cns.iu.edu/

amccrani@indiana.edu | Logout
Sci² Tool
A Tool for Science of Science Research & Practice
Home
Download
Documentation
Ask An Expert
Testimonials
Developers
The Science of Science (Sci²) Tool is a modular toolset specifically designed for the study of science. It supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels.
Registration required.
Download Sci² Tool
Mapping Topics and Topic Bursts in PNAS
3 / 4
Start Stop (4)
News
2012
Jul 24, The latest Sci² Tutorial for the National Institutes of Health, Portfolio Analysis Symposium is now available.
Jun 13, The Sci² (Science of Science) Tool v1.0 alpha release is now available, supports Mac OS X Cocoa frameworks platform and featuring new visualizations, Google Scholar Citation reader, R
Have a question?
Ask an Expert!

A science of science tool


Sci2 Tool
File Data Preparation Preprocessing Analysis Modeling Visualization R Help
Console
Welcome to the Science of Science Tool (Sci2).
The development of this tool is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0715303, and the James S. McDonnell Foundation. See the Science of Science homepage (http://sci2.wiki.cns.iu.edu) for documentation and screenshots. Please visit https://sci2.cns.iu.edu/user/ask.php if you need help with your analyses, have questions about datasets, or would like to suggest enhancements and new features.
Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc. The Sci2 tool was developed by Chin Hua Kong, Joseph Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, and Russell J. Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University. Many algorithm plugins were derived from the Network Workbench Tool (http://nwb.cns.iu.edu).
Please cite as follows:
Data Manager
CSV file: C:\Users\amccrani\AppData\Local\Temp\temp\P
Google Citation User Id
Citation Table: 0TJZPj0AAAAJ
Citation Table: 25_H0_wAAAAJ
Citation Table: U3CXAPsAAAAJ
Citation Table: RfNhTAMAAAAJ
Citation Table: 13K1czwAAAAJ
Citation Table: Fqw9o84AAAAJ
Scheduler
Remove From List
Remove completed automatically
Remove all completed
Algorithm Name
Date
Time
% Complete
Attach Citation Table fr...
05/18/2013
03:21:56 PM
Attach Citation BibTex F...
05/18/2013
02:58:13 PM
Google Citation User ID ...
05/18/2013
02:51:37 PM
Load...
05/18/2013
02:51:25 PM
Load...
05/18/2013
02:48:33 PM

Cyberinfrastructure Shell (CIShell)
CIShell supports the plug-and-play of datasets and algorithms and their bundling into custom tools that serve the specific needs of a user group or research community. It has been applied to develop diverse custom tools, see below. Feel free to take plugins from any of these tools to design your personal dream tool.

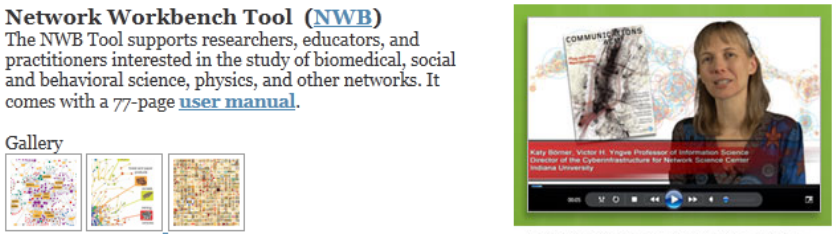
Visit the CIShell wiki
to learn more about using CIShell as a platform for your tool!

Provided by the Cyberinfrastructure for Network Science Center at Indiana University.

Learn more about existing CIShell-powered tools below.

Network Workbench Tool (NWB)
The NWB Tool supports researchers, educators, and practitioners interested in the study of biomedical, social and behavioral science, physics, and other networks. It comes with a 77-page user manual.

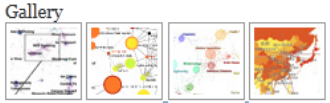
Gallery



Watch the movie about CIShell-Powered tools on the SciVee: Making Science Visible website by clicking on the image above.

Science of Science Tool (Sci²)
The Sci² Tool was specifically developed for science policy makers and researchers that study science by scientific means. It supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels. There exists a 112-page user manual, a continuously updated Sci² Tool wiki, and 24 hours of NIH tutorials in this tool.

Gallery



Epidemics Tool (EpiC)
The EpiC Tool supports the custom analysis, modeling, and visualization of data streams such as diffusion patterns of the H1N1 virus over geographic space.

Gallery

Types and level of Analysis

	Micro/Individual *(1-100 records)*	*Meso/Local* *(101–10,000 records)*	*Macro/Global* *(10,000 < records)*
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of USA, all of science.
Temporal Analysis (When)	Funding portfolio of one individual	Mapping topic bursts in 20-years of PNAS	113 Years of physics Research
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a states intellectual landscape	PNAS publications
Topical Analysis (What)	Base knowledge from which one grant draws.	Knowledge flows in Chemistry research	VxOrd/Topic maps of NIH funding
Network Analysis (With Whom?)	NSF Co-PI network of one individual	Co-author network	NSF's core competency

Why (or why not) to use Sci2

Pros

- Friendly interface
- Strong, extensive documentation with sample workflows
- Powerful parsing capabilities for both general and specific file types
- Works well in using and producing common data products so movement between programs is not difficult (R, Pajek, visone, Gephi)
- Can handle very large datasets
- Can be customized with plug-ins

Cons

- Fewer (ready) implemented algorithms than statnet, UCINET, or Pajek
- Less flexible visualization within the program
- Adding plug-ins can be daunting.

For more information about other software that could be of interest for the study of science, see section 8.2 of the Sci^2 user manual for listing and discussion.

Data suited for this tool

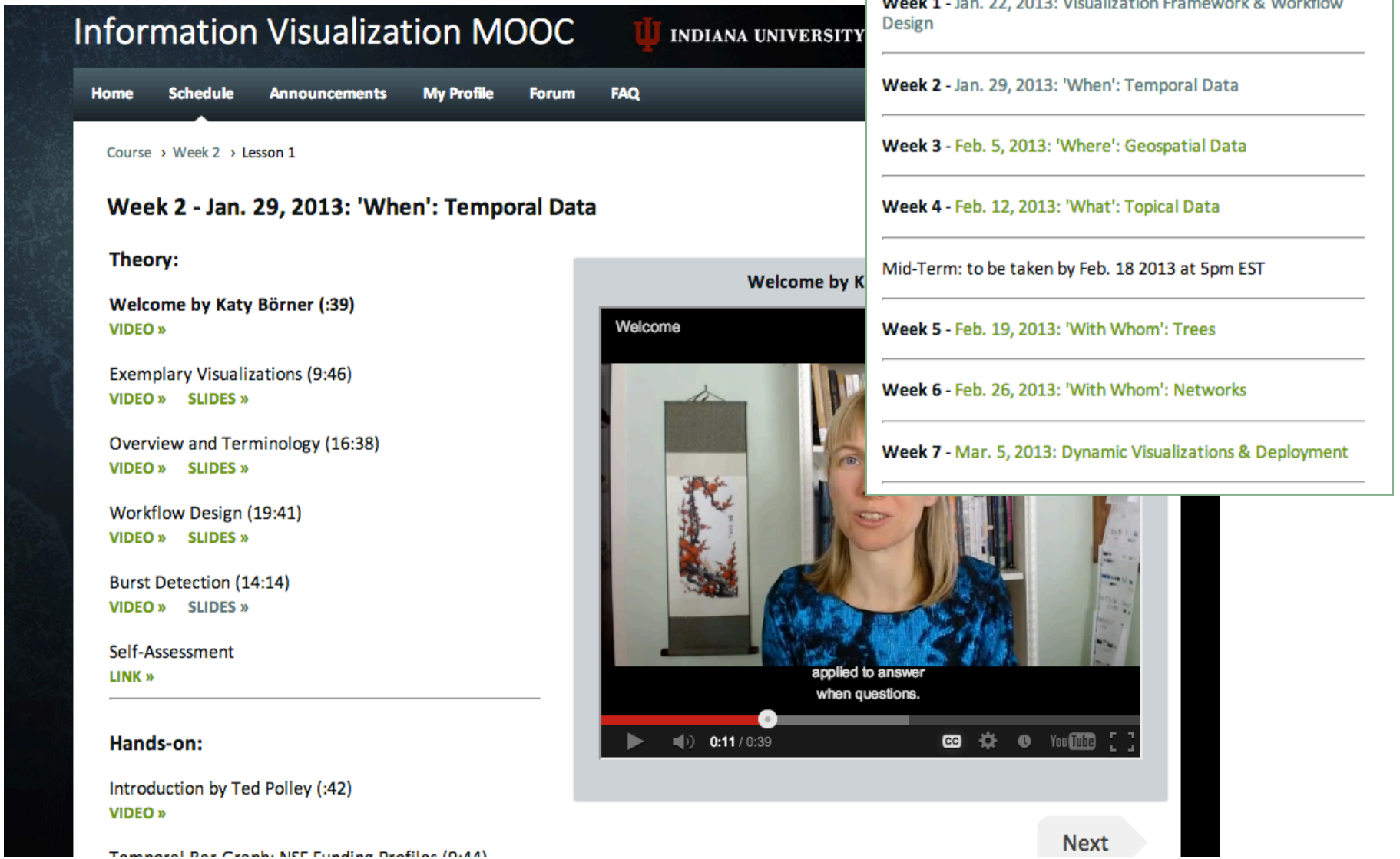
- Publications
  - Refer/BibIX/enw
  - BibTeX
  - ISI Web of Science
  - Scopus
  - Google Scholar
  - Google Citation
- Funding
  - NSF Award Search
  - NIH RePORTER
  - Scholarly Database

Specific file formats

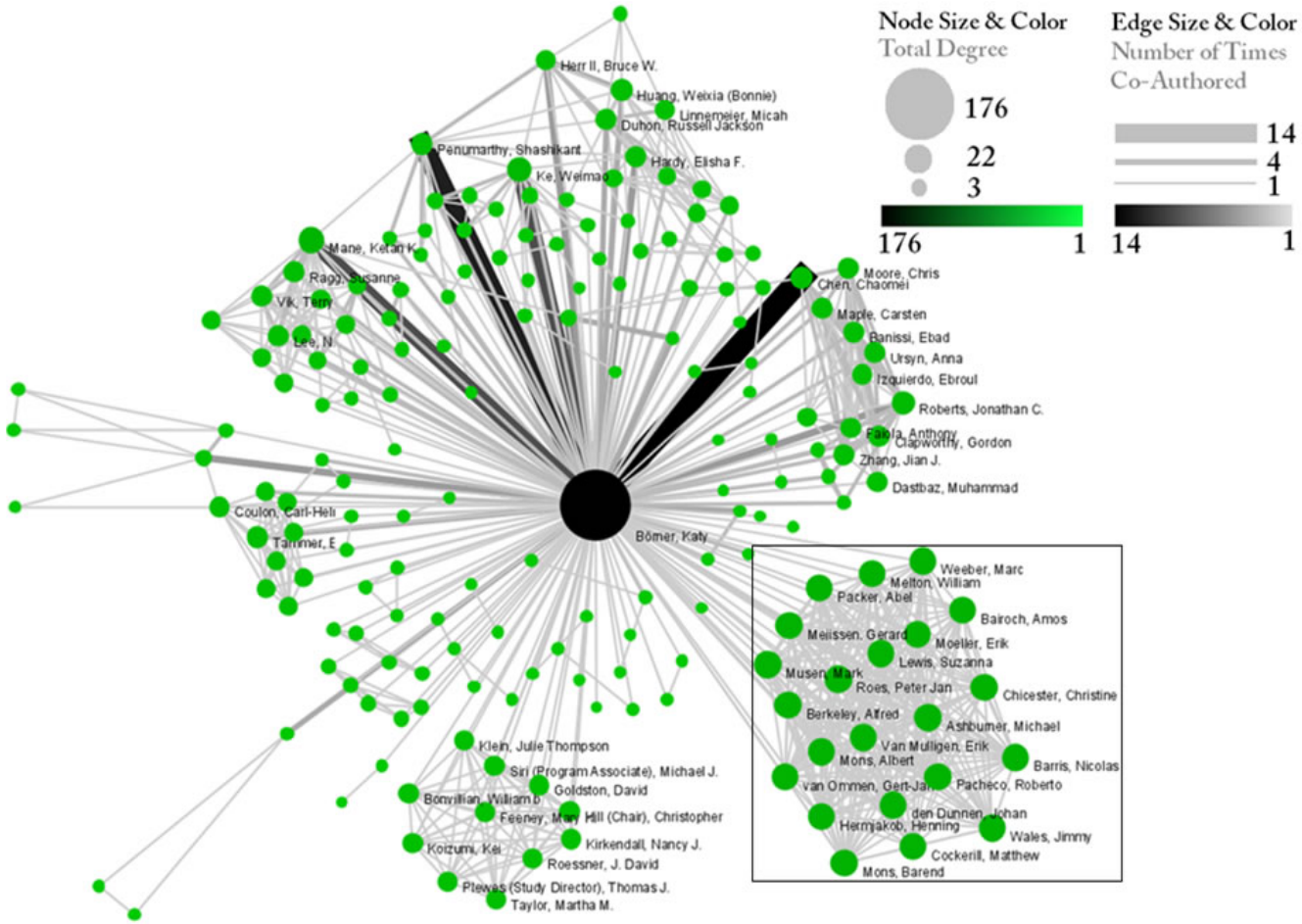
- Network Formats
  - GraphML (*.xml or *.graphml)
  - XGMML (*.xml)
  - Pajek .NET (*.net)
  - NWB (*.nwb)
  - Scientometric Formats
  - ISI (*.isi)
  - Bibtex (*.bib)
  - Endnote Export Format (*.enw)
  - Scopus csv (*.scopus)
  - NSF csv (*.nsf)
- Other Formats
  - Pajek Matrix (*.mat)
  - TreeML (*.xml)
  - Edgelist (*.edge)
  - CSV (*.csv)
- Databases
- Scholarly Database: http://sdb.cns.iu.edu/search/

Free Online Course:

http://ivmooc.appspot.com/course


Information Visualization MOOC
INDIANA UNIVERSITY
Home
Schedule
Announcements
My Profile
Forum
FAQ
Course › Week 2 › Lesson 1
Week 2 - Jan. 29, 2013: 'When': Temporal Data
Theory:
Welcome by Katy Börner (:39)
VIDEO »
Exemplary Visualizations (9:46)
VIDEO » SLIDES »
Overview and Terminology (16:38)
VIDEO » SLIDES »
Workflow Design (19:41)
VIDEO » SLIDES »
Burst Detection (14:14)
VIDEO » SLIDES »
Self-Assessment
LINK »
Hands-on:
Introduction by Ted Polley (:42)
VIDEO »
Welcome by K
Welcome
applied to answer
when questions.
0:11 / 0:39
Next
Schedule
Pre-Questionnaire
Week 1 - Jan. 22, 2013: Visualization Framework & Workflow Design
Week 2 - Jan. 29, 2013: 'When': Temporal Data
Week 3 - Feb. 5, 2013: 'Where': Geospatial Data
Week 4 - Feb. 12, 2013: 'What': Topical Data
Mid-Term: to be taken by Feb. 18 2013 at 5pm EST
Week 5 - Feb. 19, 2013: 'With Whom': Trees
Week 6 - Feb. 26, 2013: 'With Whom': Networks
Week 7 - Mar. 5, 2013: Dynamic Visualizations & Deployment

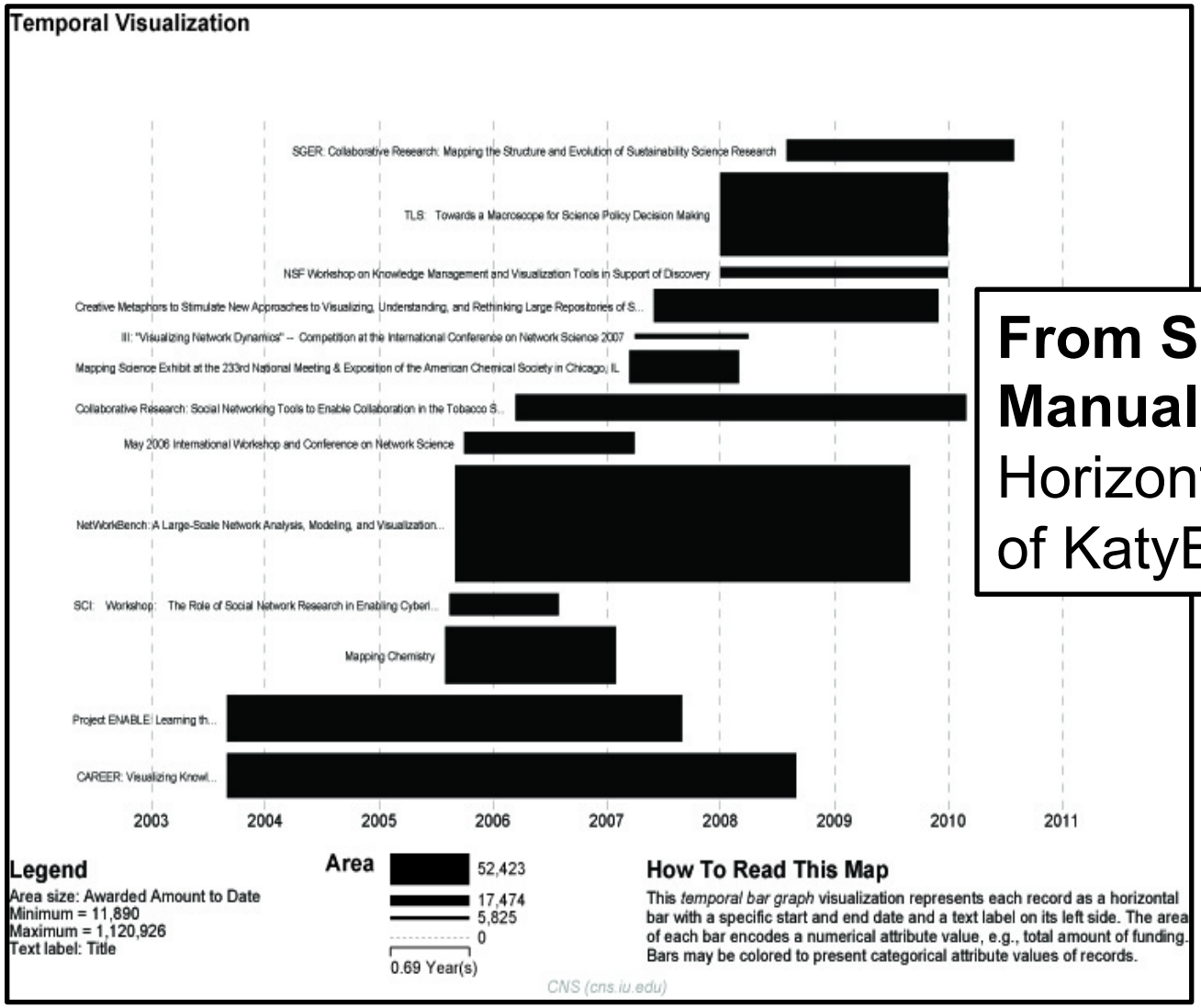
Micro: Individual Scientist



From Sci² User Manual: Figure 5.1: Co-authorship network of Katy Börner
GUESS supports the repositioning of selected nodes. Multiple nodes can be selected by holding down the 'Shift' key and dragging a box around specific nodes. The final network can be saved via 'GUESS: File > Export Image' and opened in a graphic design program to add a title and legend. The image above was created using Adobe Photoshop. Node clusters were highlighted and increased in size, the label font size was increased for emphasis, and a legend was added to clarify the significance of node and edge size and color.

http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/5.1+Individual+Level+Studies+-+Micro

Micro: Trends for Individual Scientist


Temporal Visualization
SGER: Collaborative Research: Mapping the Structure and Evolution of Sustainability Science Research
TLS: Towards a Macroscope for Science Policy Decision Making
NSF Workshop on Knowledge Management and Visualization Tools in Support of Discovery
Creative Metaphors to Stimulate New Approaches to Visualizing, Understanding, and Rethinking Large Repositories of S...
III: "Visualizing Network Dynamics" -- Competition at the International Conference on Network Science 2007
Mapping Science Exhibit at the 233rd National Meeting & Exposition of the American Chemical Society in Chicago, IL
Collaborative Research: Social Networking Tools to Enable Collaboration in the Tobacco S...
May 2006 International Workshop and Conference on Network Science
NetWorkBench: A Large-Scale Network Analysis, Modeling, and Visualization...
SCI: Workshop: The Role of Social Network Research in Enabling Cyberi...
Mapping Chemistry
Project ENABLE: Learning th...
CAREER: Visualizing Knowl...
2003
2004
2005
2006
2007
2008
2009
2010
2011
Legend
Area size: Awarded Amount to Date
Minimum = 11,890
Maximum = 1,120,926
Text label: Title
Area
52,423
17,474
5,825
0
0.69 Year(s)
How To Read This Map
This temporal bar graph visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes a numerical attribute value, e.g., total amount of funding. Bars may be colored to present categorical attribute values of records.
CNS (cns.iu.edu)

From Sci² User Manual: Figure 5.3: Horizontal Bar Graph of KatyBorner.nsf

http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/5.1+Individual+Level+Studies+-+Micro

Micro: Trends for Individual Scientist

Alessandro Vespignani

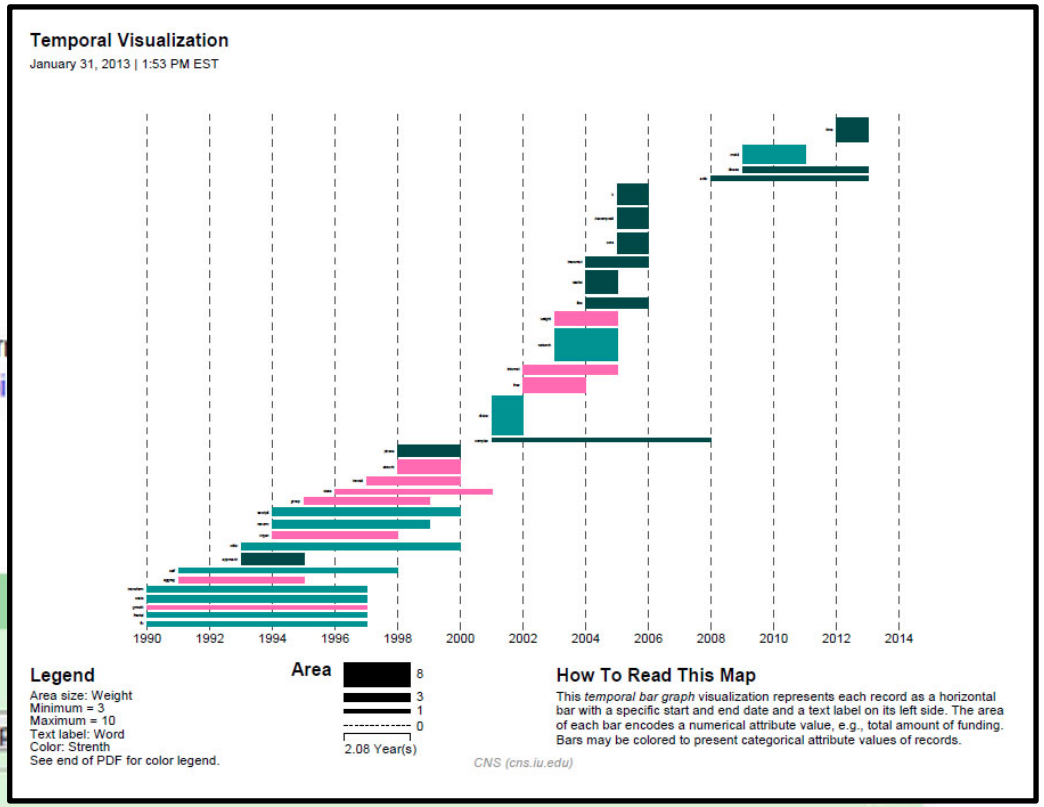
Sternberg Distinguished Professor, Northeastern

computational sciences - Network Science - Epi

Physics

Verified email at neu.edu

Homepage



Export articles

File format: BibTeX

BibTeX
EndNote
RefMan
CSV

Export the ... elow

Export all articles by Alessandro Vesp

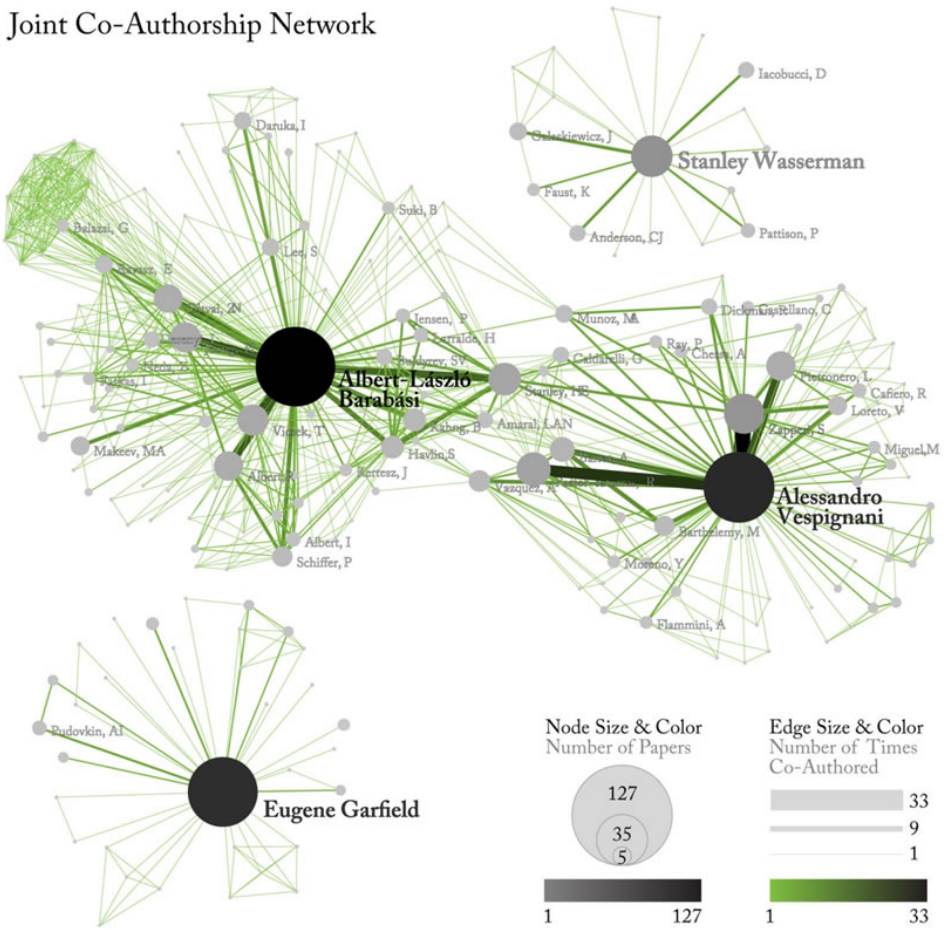
Title / Author

Epidemic spreading in scale-free networks R Pastor-Satorras, A Vespignani Physical review letters 86 (14), 3200-3203	2001
The architecture of complex weighted networks A Barrat, M Barthelemy, R Pastor-Satorras, A Vespignani Proceedings of the National Academy of Sciences of the United States of ...	2004

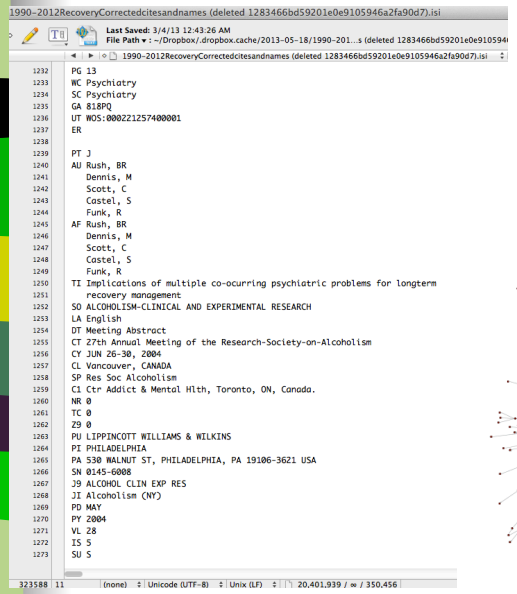
From Sci² User Manual:

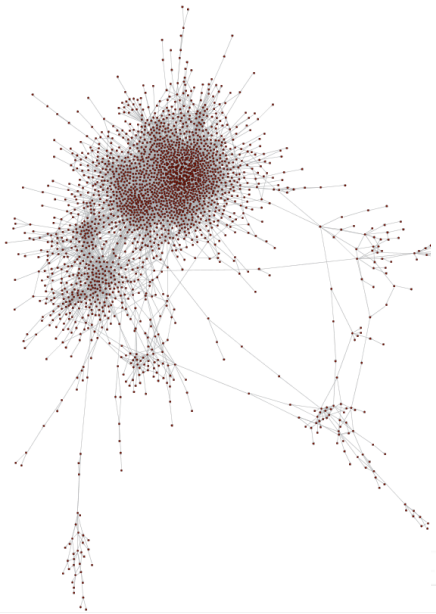
http://sci2.wiki.cns.iu.edu/display/SCI2TUTORIAL/5.2+Institution+Level+Studies+-+Meso

Micro: Four Scientists

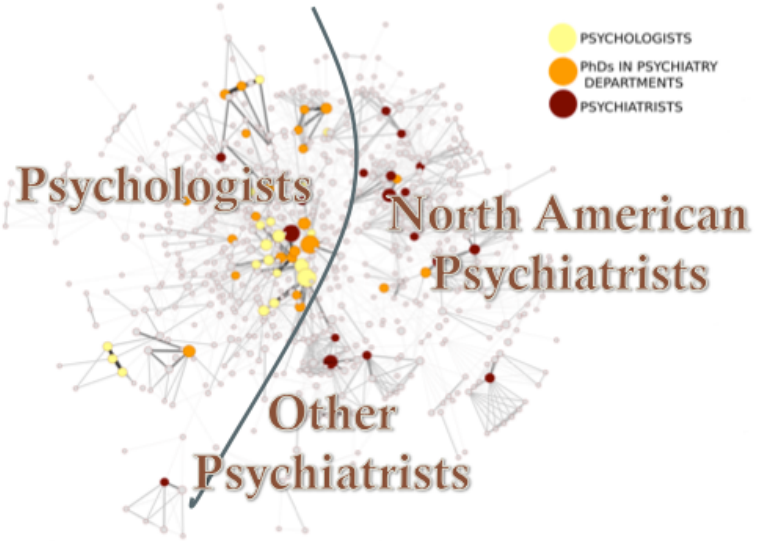


Meso: A Research Field





From: McCranie 2013, unpublished dissertation in progress. Image in middle above is largest single component of the paper-citation network of “recovery” research literature from 1991-2012 . Top right image is Pathfinder Network Scaling of the most cited works in transformed co-citation network color coded by content analysis of articles. Bottom right is co-authorship network coded by disciplinary field. Data prep and analysis in Sci2, visualization in visone with added legends and graphic elements in Inkscape.

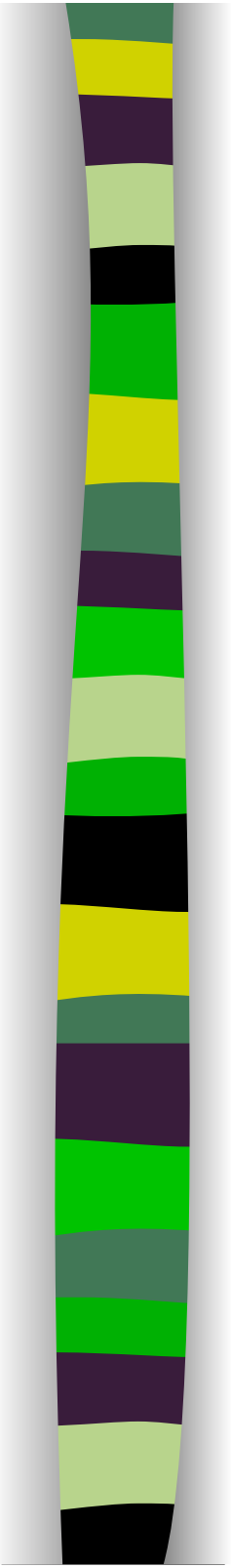


Simple, but important!

Most highly cited works	Citation Count	In ISI Dataset?
American Psychiatric Assoc, 1994, DIAGN STAT MAN MENT	406	book
Anthony WA, 1993, PSYCHOSOC REHABIL, V16, P11	257	yes
Hamilton M, 1960, J NEUROL NEUROSUR PS, V23, P56	177	too early
American Psychiatric Assoc, 1987, DIAGN STAT MAN MENT	164	too early
Kay SR, 1987, SCHIZOPHRENIA BULL, V13, P261	159	too early
Deegan PE, 1988, PSYCHOSOC REHABIL, V11, P11	155	too early
American Psychiatric Assoc, 2000, DIAGN STAT MAN MENT	144	book
Davidson L, 1992, BRIT J MED PSYCHOL, V65, P131	115	yes
Jacobson N, 2001, PSYCHIATR SERV, V52, P482	113	yes
Frank E, 1991, ARCH GEN PSYCHIAT, V48, P851	111	yes
Overall JE, 1962, PSYCHOL REP, V10, P799	99	too early
Andreasen NC, 2005, AM J PSYCHIAT, V162, P441	98	yes
Mueser KT, 2002, PSYCHIATR SERV, V53, P1272	95	yes
Liberman RP, 2002, INT REV PSYCHIATR, V14, P256	93	yes
Folstein MF, 1975, J PSYCHIAT RES, V12, P189	91	too early
Keller MB, 1987, ARCH GEN PSYCHIAT, V44, P540	89	too early
Young SL, 1999, PSYCHIATR REHABIL J, V22, P219	88	yes
Beck AT, 1961, ARCH GEN PSYCHIAT, V4, P561	88	too early
Harrison G, 2001, BRIT J PSYCHIAT, V178, P506	87	yes
Harding CM, 1987, AM J PSYCHIAT, V144, P727	85	too early
Deegan P, 1996, PSYCHIATR REHABIL J, V19, P91	83	yes
Endicott J, 1976, ARCH GEN PSYCHIAT, V33, P766	80	too early
Ridgway P, 2001, PSYCHIATR REHABIL J, V24, P335	79	yes
Kessler RC, 1994, ARCH GEN PSYCHIAT, V51, P8	74	does not fit search

The Demonstration

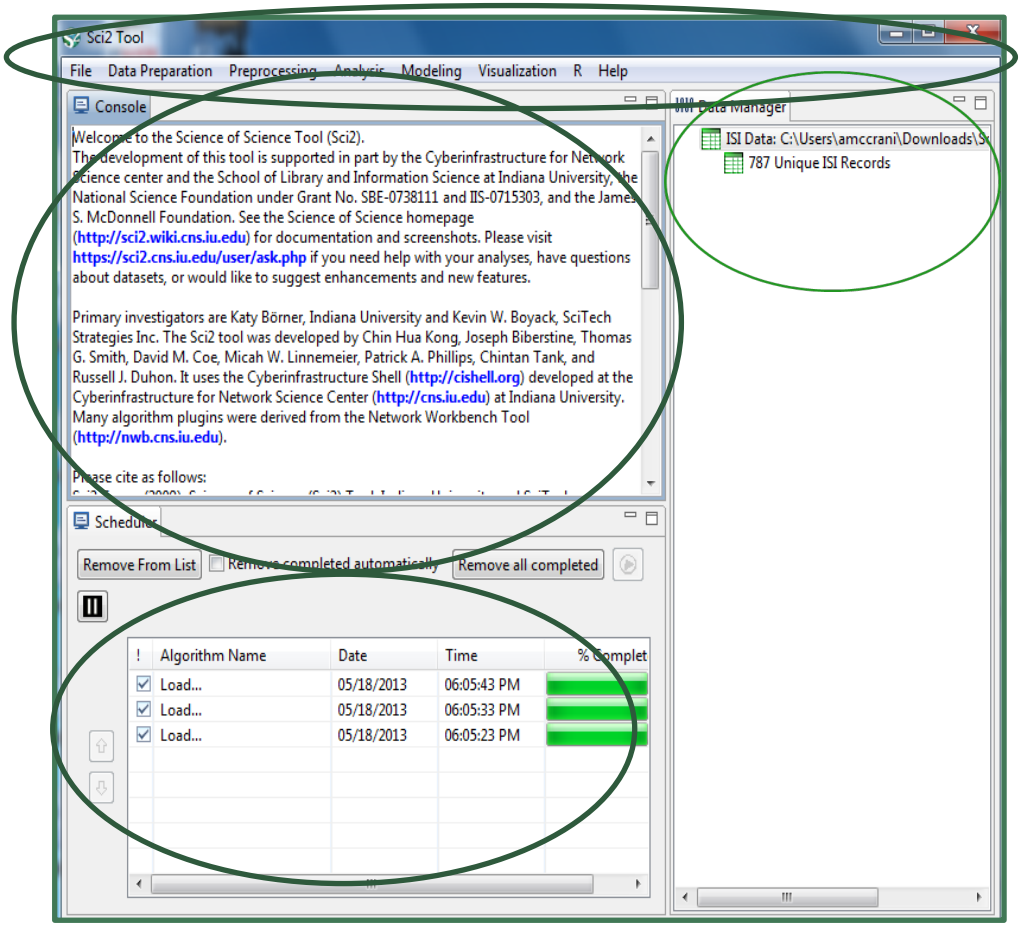
- Basic Features of Sci2
- Citation and co-authorship networks (with special emphasis on preparing and cleaning data)
- Topical analysis and visualization
- Available data and ways to pull it into Sci2
- Questions & discussion



THE BASIC FUNCTIONALITY

The Basics

- Opening Data
- Data Preparation
- Preprocessing
- Analysis
- Modeling
- Visualization


Sci2 Tool
File
Data Preparation
Preprocessing
Analysis
Modeling
Visualization
R
Help
Console
Welcome to the Science of Science Tool (Sci2).
The development of this tool is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0715303, and the James S. McDonnell Foundation. See the Science of Science homepage (http://sci2.wiki.cns.iu.edu) for documentation and screenshots. Please visit https://sci2.cns.iu.edu/user/ask.php if you need help with your analyses, have questions about datasets, or would like to suggest enhancements and new features.
Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc. The Sci2 tool was developed by Chin Hua Kong, Joseph Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, and Russell J. Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University. Many algorithm plugins were derived from the Network Workbench Tool (http://nwb.cns.iu.edu).
Please cite as follows:
Scheduler
Remove From List
Remove completed automatically
Remove all completed
Algorithm Name
Date
Time
% Complet
Load...
05/18/2013
06:05:43 PM
Load...
05/18/2013
06:05:33 PM
Load...
05/18/2013
06:05:23 PM
Data Manager
ISI Data: C:\Users\amccrani\Downloads\S
787 Unique ISI Records

Visualizing the Florentine dataset

- Data Acquisition & Preprocessing
  - Examine the data
  - Load the data into NWB
- Data Analysis, Modeling, & Layout
  - Since the data is formatted as a network with various attributes added to the nodes and edges – only simple analysis will be conducted on this network.
- Data communication & Visualization Layers
  - Visualize the Florentine network with GUESS

Visualizing the Florentine dataset

- Florentine families related through business ties (specifically, recorded financial ties such as loans, credits and joint partnerships) and marriage alliances.

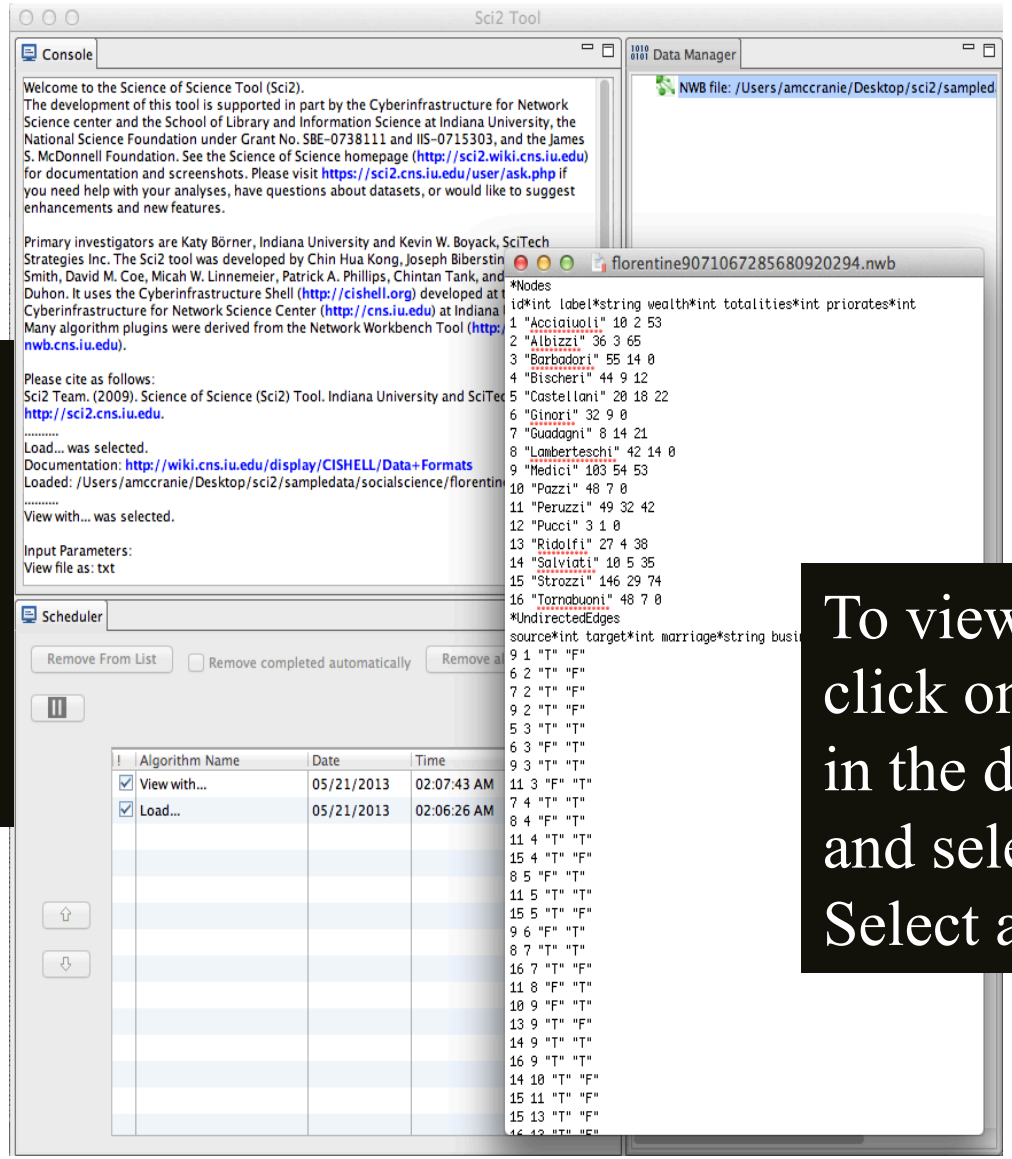
- Node attributes
  - Wealth: Each family's net wealth in 1427 (in thousands of lira).
  - Priorates: The number of seats on the civic council held between 1282-1344.
  - Totalities: Number of business/marriage ties in complete dataset of 116 families.

- Edge attributes: Marriage T/F & Business T/F

- “Substantively, the data include families who were locked in a struggle for political control of the city of Florence around 1430. Two factions were dominant in this struggle: one revolved around the infamous Medicis, the other around the powerful Strozzis.”
- More info is at http://svitsrv25.epfl.ch/R-doc/library/ergm/html/florentine.html and Padgett & Ansell 1993 (http://home.uchicago.edu/~jpadgett/papers/published/robust.pdf)

Visualizing the Florentine dataset

FILE> LOAD the Florentine dataset: sampledata/ socialscience/ florentine.nwb


Sci2 Tool
Console
Welcome to the Science of Science Tool (Sci2).
The development of this tool is supported in part by the Cyberinfrastructure for Network Science center and the School of Library and Information Science at Indiana University, the National Science Foundation under Grant No. SBE-0738111 and IIS-0715303, and the James S. McDonnell Foundation. See the Science of Science homepage (http://sci2.wiki.cns.iu.edu) for documentation and screenshots. Please visit https://sci2.cns.iu.edu/user/ask.php if you need help with your analyses, have questions about datasets, or would like to suggest enhancements and new features.
Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc. The Sci2 tool was developed by Chin Hua Kong, Joseph Biberstin
Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, and
Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org) developed at t
Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana
Many algorithm plugins were derived from the Network Workbench Tool (http:/
nwb.cns.iu.edu).
Please cite as follows:
Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTec
http://sci2.cns.iu.edu.
Load... was selected.
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Data+Formats
Loaded: /Users/amccranie/Desktop/sci2/sampledata/socialscience/florentin
View with... was selected.
Input Parameters:
View file as: txt
Data Manager
NWB file: /Users/amccranie/Desktop/sci2/sampled
florentine9071067285680920294.nwb
*Nodes
id*int label*string wealth*int totalities*int priorates*int
1 "Acciaiuoli" 10 2 53
2 "Albizzi" 36 3 65
3 "Barbadori" 55 14 0
4 "Bischeri" 44 9 12
5 "Castellani" 20 18 22
6 "Ginori" 32 9 0
7 "Guadagni" 8 14 21
8 "Lamberteschi" 42 14 0
9 "Medici" 103 54 53
10 "Pazzi" 48 7 0
11 "Peruzzi" 49 32 42
12 "Pucci" 3 1 0
13 "Ridolfi" 27 4 38
14 "Salviati" 10 5 35
15 "Strozzi" 146 29 74
16 "Tornabuoni" 48 7 0
*UndirectedEdges
source*int target*int marriage*string busi
9 1 "T" "F"
6 2 "T" "F"
7 2 "T" "F"
9 2 "T" "F"
5 3 "T" "T"
6 3 "F" "T"
9 3 "T" "T"
11 3 "F" "T"
7 4 "T" "T"
8 4 "F" "T"
11 4 "T" "T"
15 4 "T" "F"
8 5 "F" "T"
11 5 "T" "T"
15 5 "T" "F"
9 6 "F" "T"
8 7 "T" "T"
16 7 "T" "F"
11 8 "F" "T"
10 9 "F" "T"
13 9 "T" "F"
14 9 "T" "T"
16 9 "T" "T"
14 10 "T" "F"
15 11 "T" "F"
15 13 "T" "F"
Scheduler
Remove From List
Remove completed automatically
Algorithm Name
Date
Time
View with...
05/21/2013
02:07:43 AM
Load...
05/21/2013
02:06:26 AM

To view the file, right click on the network in the data manager and select view as. Select a text editor.

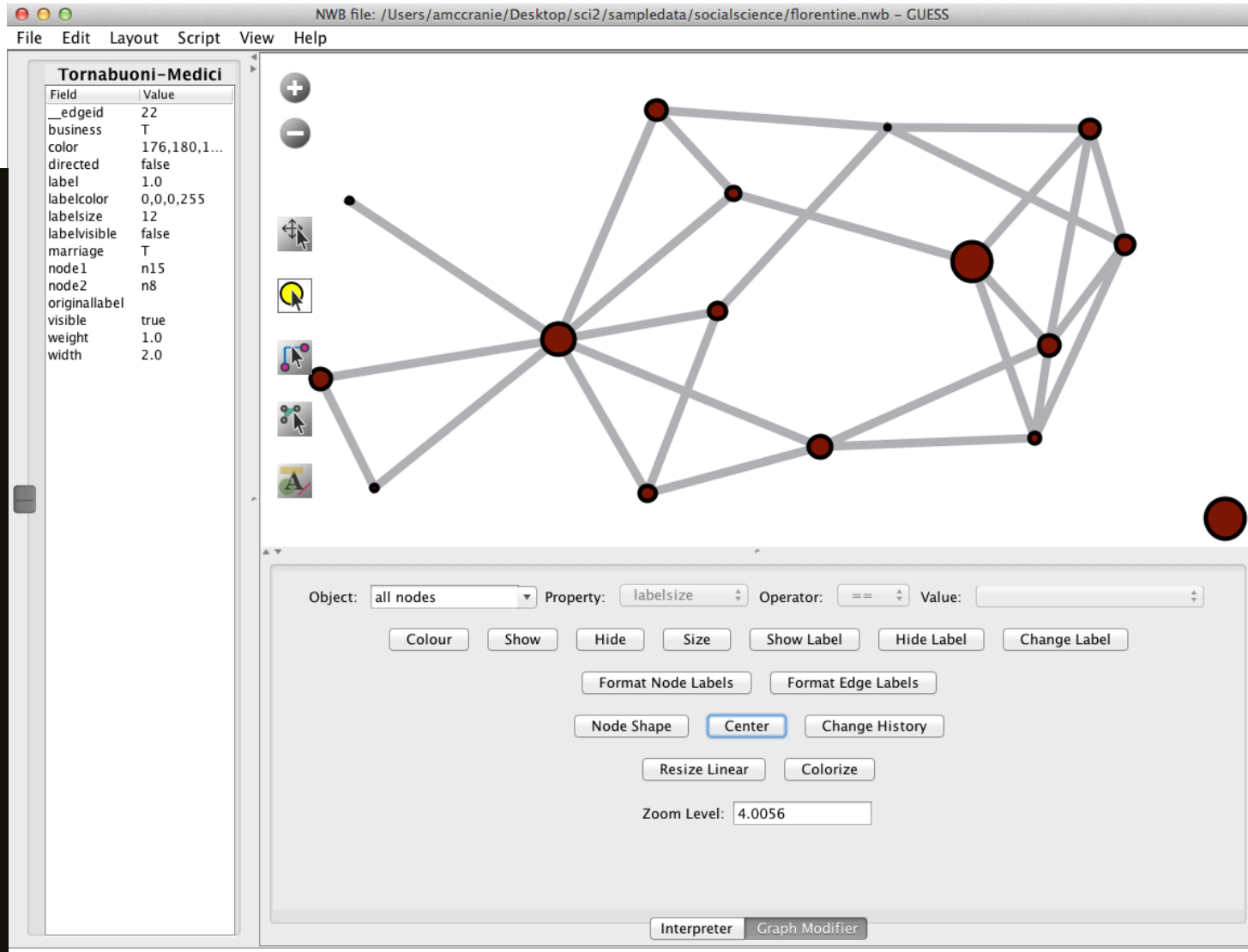
Visualizing the Florentine dataset

```
*Nodes
id*int label*string wealth*int totalities*int priorates*int
1 "Acciaiuoli" 10 2 53
2 "Albizzi" 36 3 65
3 "Barbadori" 55 14 0
4 "Bischeri" 44 9 12
5 "Castellani" 20 18 22
6 "Ginori" 32 9 0
7 "Guadagni" 8 14 21
8 "Lamberteschi" 42 14 0
9 "Medici" 103 54 53
10 "Pazzi" 48 7 0
11 "Peruzzi" 49 32 42
12 "Pucci" 3 1 0
13 "Ridolfi" 27 4 38
14 "Salviati" 10 5 35
15 "Strozzi" 146 29 74
16 "Tornabuoni" 48 7 0
*UndirectedEdges
source*int target*int marriage*string business*string
9 1 "T" "F"
6 2 "T" "F"
7 2 "T" "F"
9 2 "T" "F"
5 3 "T" "T"
```

Visualizing the Florentine dataset

To visualize the Florentine network select Visualization > GUESS and NWB will launch GUESS.

To change the layout in GUESS select Layout > GEM (you can run GEM multiple times to randomly generate a network to your satisfaction, as seen to the right).



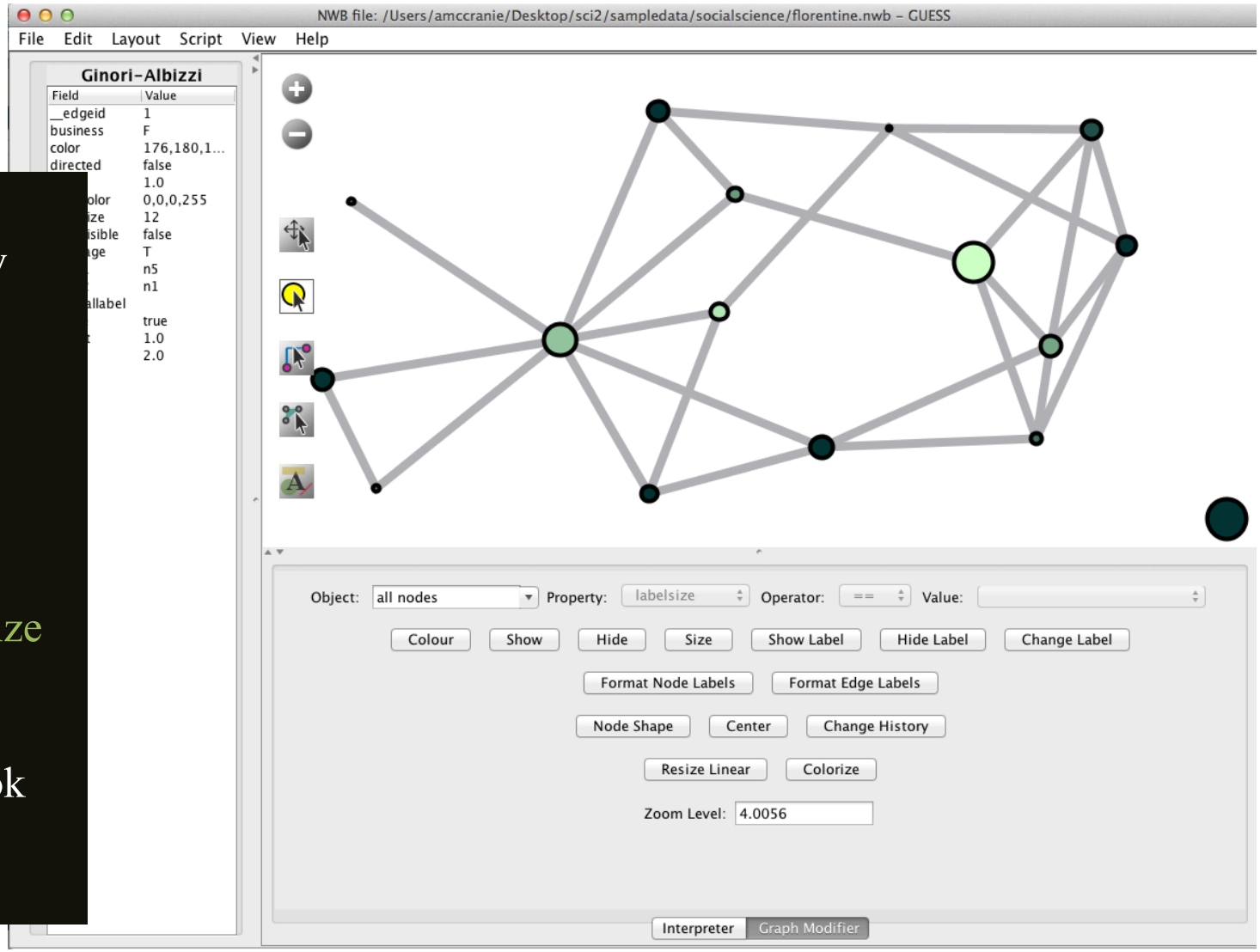
Visualizing the Florentine dataset

Resize the nodes according to family wealth.

Resize Linear >Nodes> Wealth
From: 1 To: 10

Then click Do Resize Linear

The results will look similar to what is shown to the right

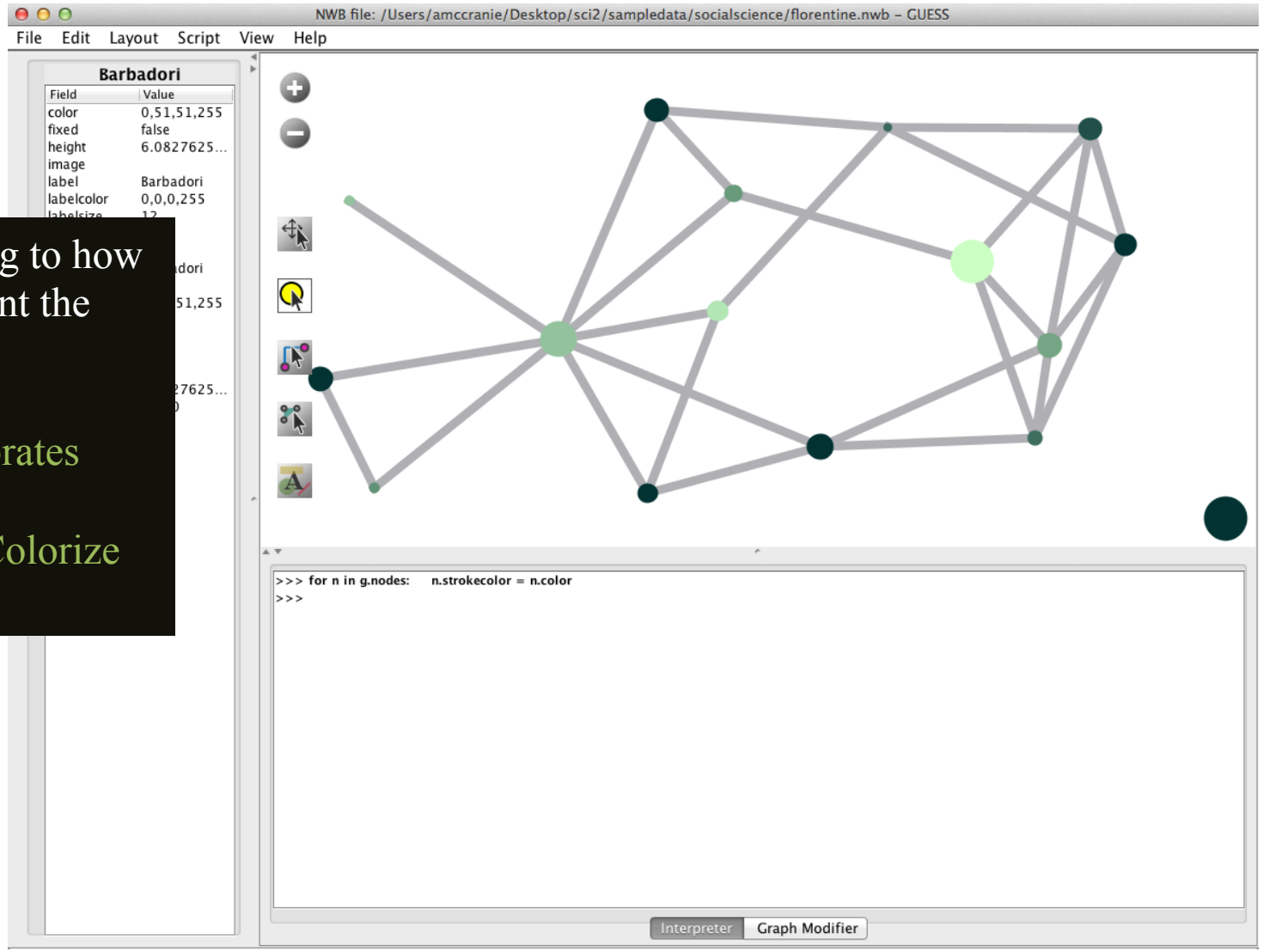


Visualizing the Florentine dataset

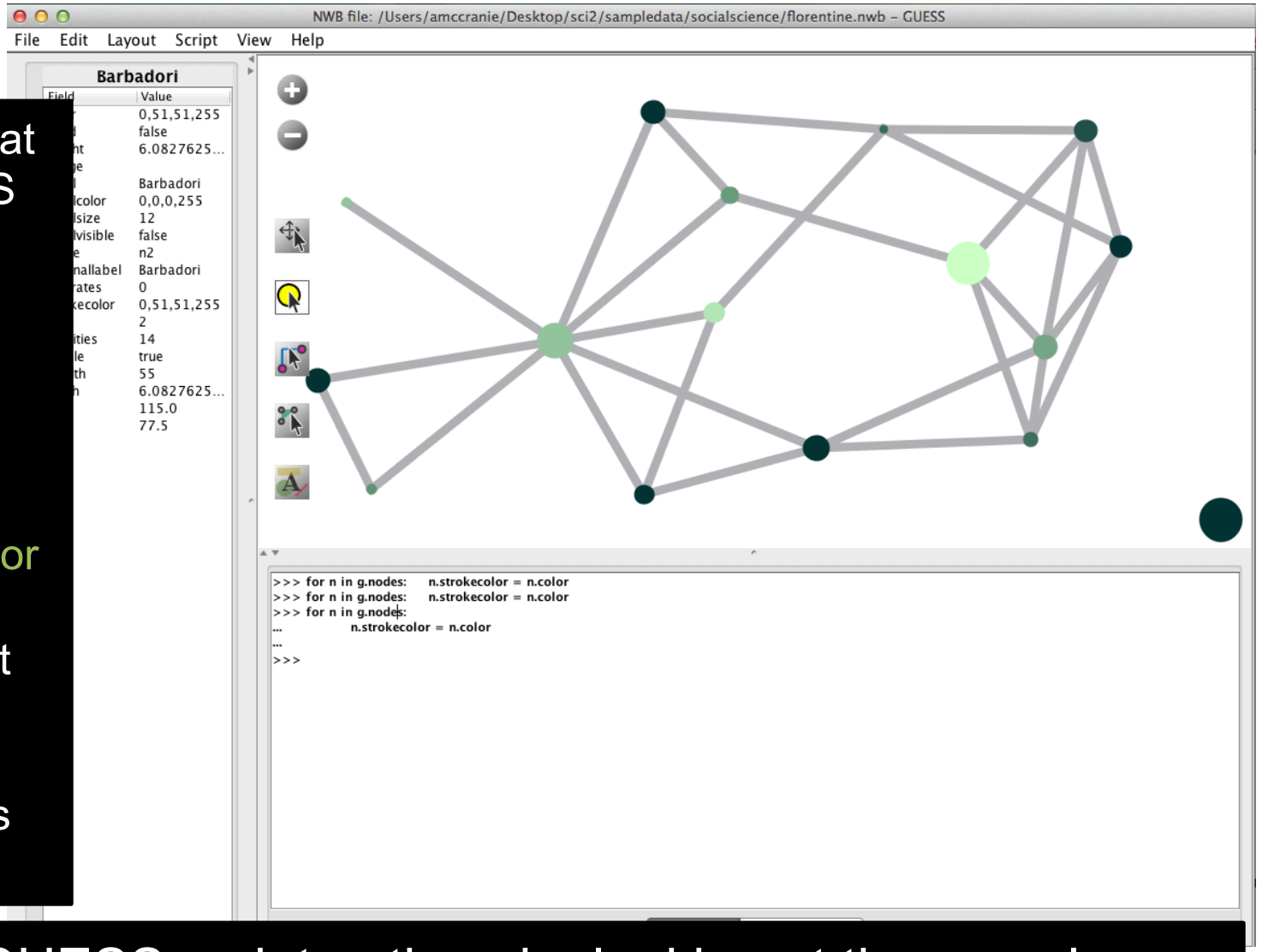
Colorize nodes according to how many seats in government the family holds.

Colorize > Nodes > Priorates

From : To: Do Colorize



Visualizing the Florentine dataset



Switch to the Interpreter at the bottom of the GUESS window.

Type the following commands:

```
for n in g.nodes:
    n.strokecolor = n.color
```

Note: after typing the first line hit the *Tab* key and after the second line hit *Enter* and the commands will be executed.

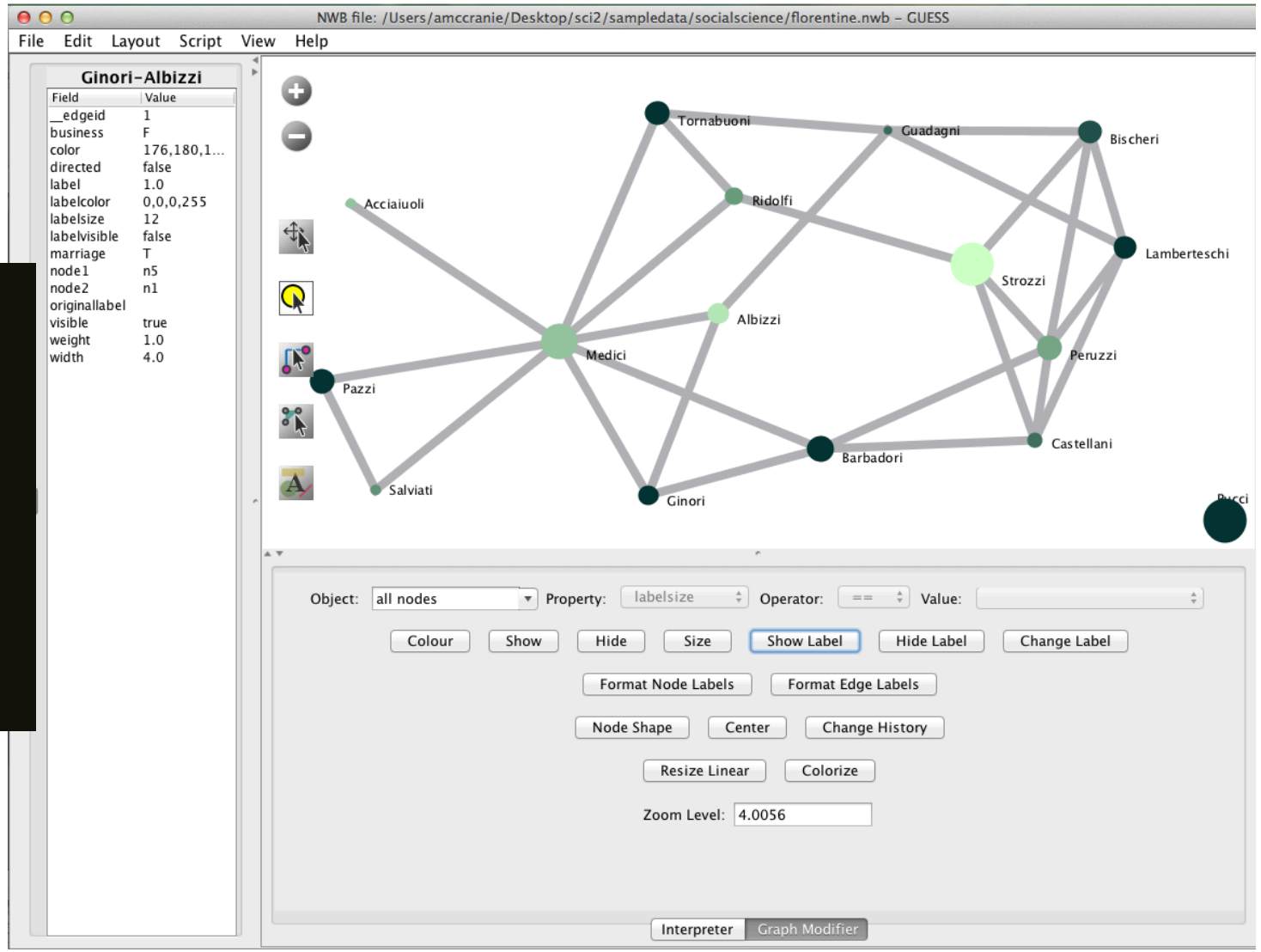
Learn more about GUESS script options by looking at the sample .py files in the sampledata folders and by visiting http://graphexploration.cond.org/documentation.html

Visualizing the Florentine dataset

Add the family labels to the nodes.

Select Object: All Nodes
Then Click Show Label

The family names will then appear next to their corresponding nodes.



Basic Analysis on the Florentine Dataset

Add betweenness centrality to the dataset.

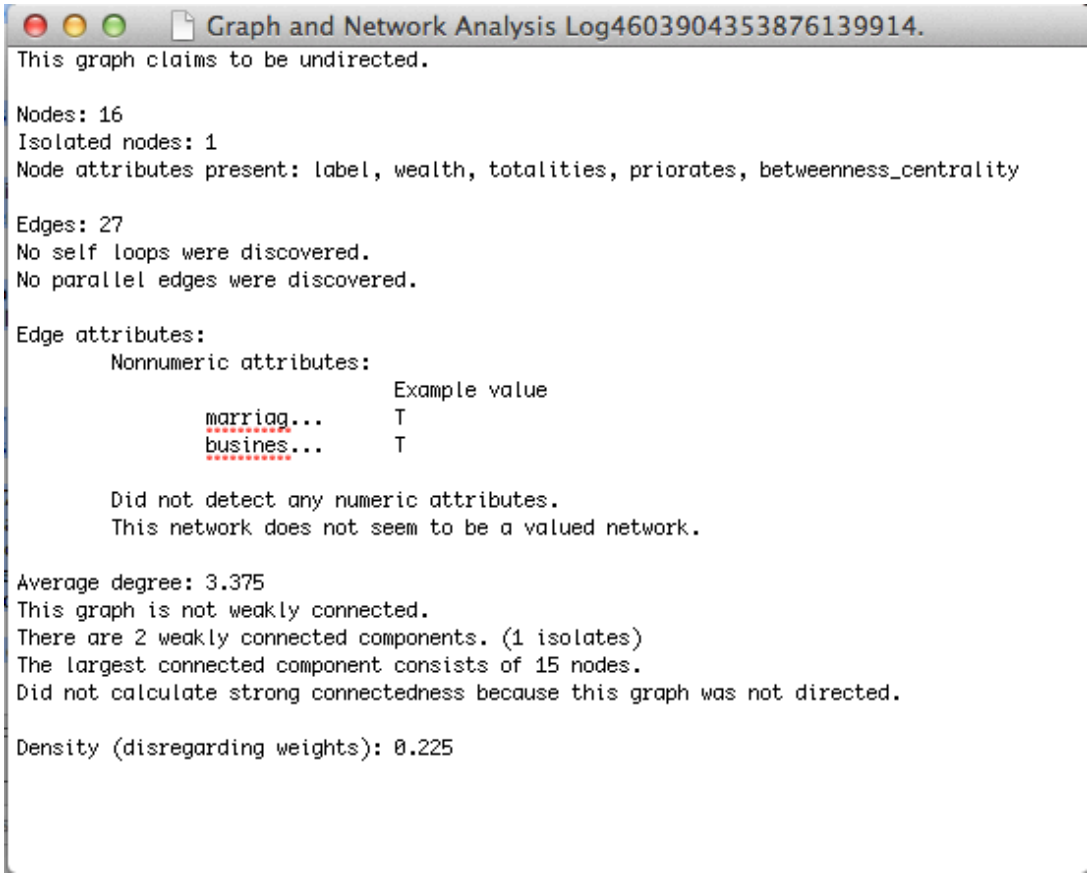
Analysis>Networks>
Unweighted & Undirected>
Node betweenness

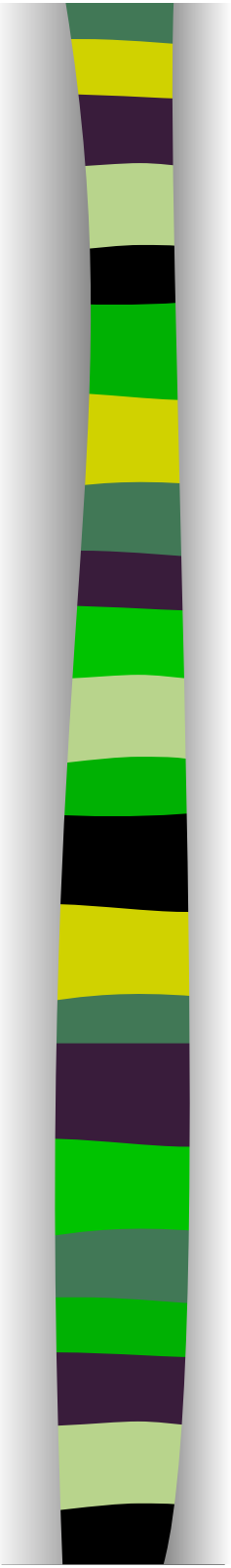
Then right click to view the new file. Remember you will have to save it if you want to use it after you have closed the program.

Now use the NAT to get basic stats on the network.

Analysis>Networks>Network Analysis Toolkit.

It will be reported in your console window and as a output log in the data manager.


Graph and Network Analysis Log4603904353876139914.
This graph claims to be undirected.
Nodes: 16
Isolated nodes: 1
Node attributes present: label, wealth, totalities, priorates, betweenness_centrality
Edges: 27
No self loops were discovered.
No parallel edges were discovered.
Edge attributes:
Nonnumeric attributes:
Example value
marriag... T
busines... T
Did not detect any numeric attributes.
This network does not seem to be a valued network.
Average degree: 3.375
This graph is not weakly connected.
There are 2 weakly connected components. (1 isolates)
The largest connected component consists of 15 nodes.
Did not calculate strong connectedness because this graph was not directed.
Density (disregarding weights): 0.225



CO-AUTHORSHIP OF A NEW JOURNAL

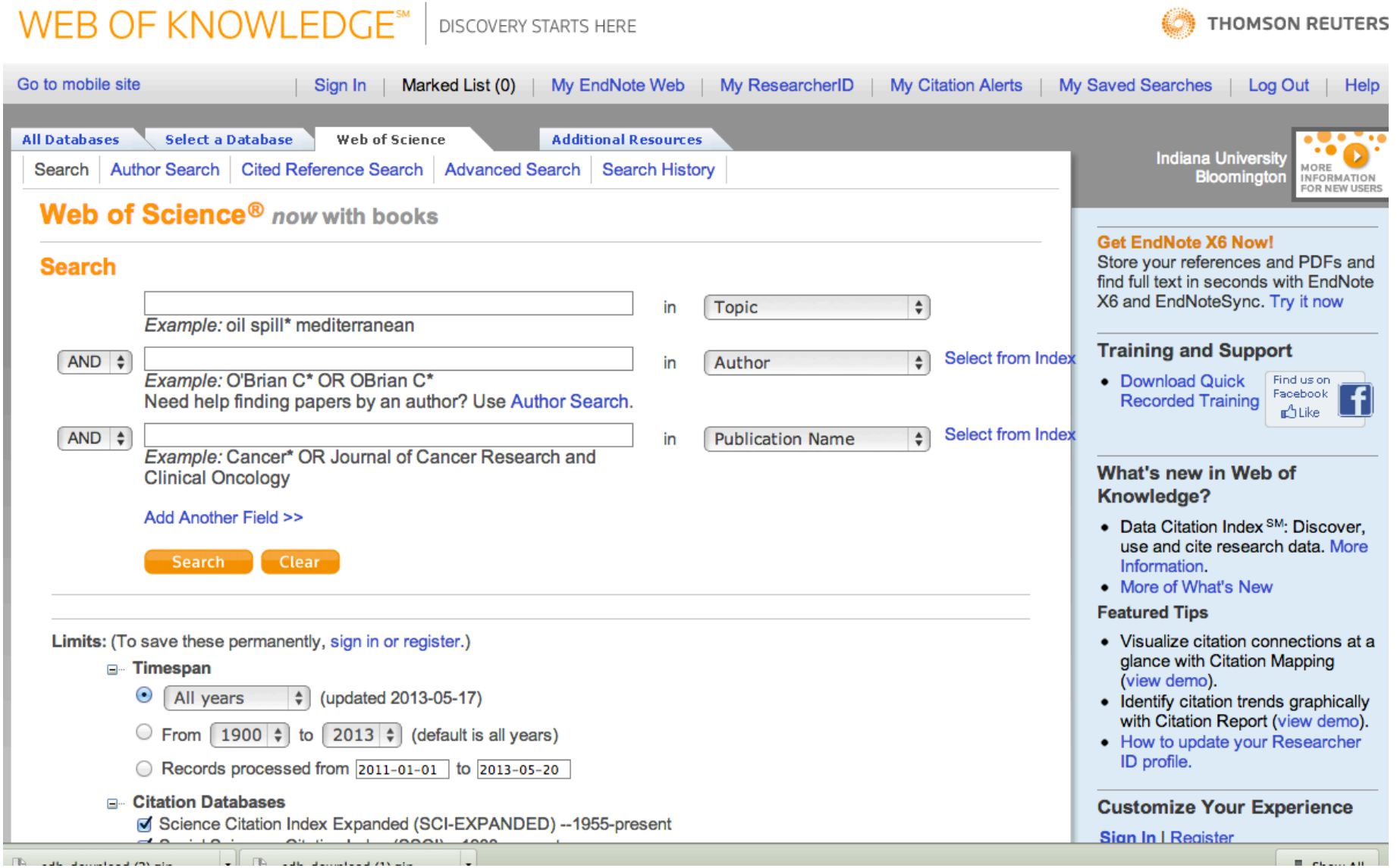
Constructing a Co-Authorship Network

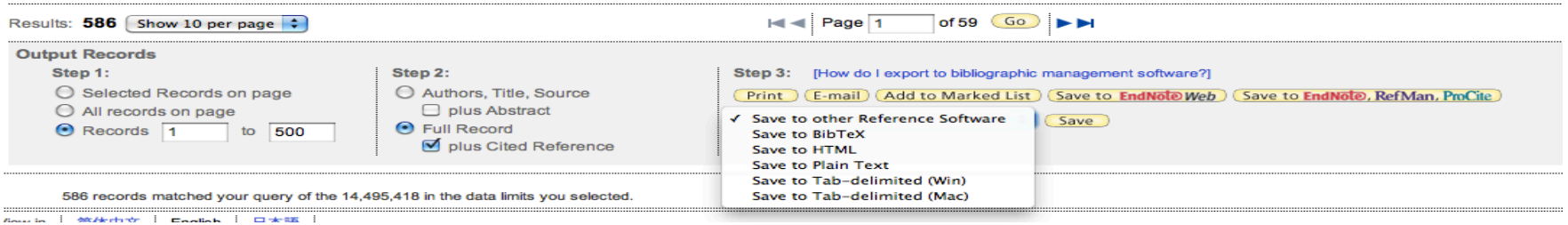
- Acquire the data (We'll use ISI, but you have many choices)
- Load and prepare the data
- CLEAN and verify the data
  - Duplicates, loops, errors, misspellings, etc…
- Analysis
- Visualization

Constructing a Co-Authorship Network

- These data were gathered by searching ISI Web of Science, a well-established bibliographic database that has wide coverage in science, social science, and the humanities. Details of each individual search are noted below. Please note that you might get slightly different results when you attempt to search, particularly if ISI Web of Science has added journals to their database in the meantime.

- Once you have conducted your search, you can export the search results into a number of formats. To create the ISI files like we we use in this this tutorial, you will need to save them as “Plain Text.” Note that you can only export 500 records at a time. If you have a search with more records, save them 500 records at a time and splice the files together, removing the notations for beginning and ending files from the second (and third, etc) set of records you add to your first file.


WEB OF KNOWLEDGE℠
DISCOVERY STARTS HERE
THOMSON REUTERS
Go to mobile site | Sign In | Marked List (0) | My EndNote Web | My ResearcherID | My Citation Alerts | My Saved Searches | Log Out | Help
All Databases | Select a Database | Web of Science | Additional Resources
Search | Author Search | Cited Reference Search | Advanced Search | Search History
Indiana University Bloomington
MORE INFORMATION FOR NEW USERS
Web of Science® now with books
Search
Example: oil spill* mediterranean
in Topic
AND
Example: O'Brian C* OR OBrian C*
Need help finding papers by an author? Use Author Search.
in Author
Select from Index
AND
Example: Cancer* OR Journal of Cancer Research and Clinical Oncology
in Publication Name
Select from Index
Add Another Field >>
Search
Clear
Limits: (To save these permanently, sign in or register.)
Timespan
All years (updated 2013-05-17)
From 1900 to 2013 (default is all years)
Records processed from 2011-01-01 to 2013-05-20
Citation Databases
Science Citation Index Expanded (SCI-EXPANDED) --1955-present
Get EndNote X6 Now!
Store your references and PDFs and find full text in seconds with EndNote X6 and EndNoteSync. Try it now
Training and Support
Download Quick Recorded Training
Find us on Facebook
Like
What's new in Web of Knowledge?
Data Citation Index℠: Discover, use and cite research data. More Information.
More of What's New
Featured Tips
Visualize citation connections at a glance with Citation Mapping (view demo).
Identify citation trends graphically with Citation Report (view demo).
How to update your Researcher ID profile.
Customize Your Experience
Sign In | Register


Results: 586 Show 10 per page
Page 1 of 59 Go
Output Records
Step 1:
Selected Records on page
All records on page
Records 1 to 500
Step 2:
Authors, Title, Source
plus Abstract
Full Record
plus Cited Reference
Step 3: [How do I export to bibliographic management software?]
Print | E-mail | Add to Marked List | Save to EndNote Web | Save to EndNote, RefMan, ProCite
Save to other Reference Software
Save to BibTeX
Save to HTML
Save to Plain Text
Save to Tab-delimited (Win)
Save to Tab-delimited (Mac)
Save
586 records matched your query of the 14,495,418 in the data limits you selected.

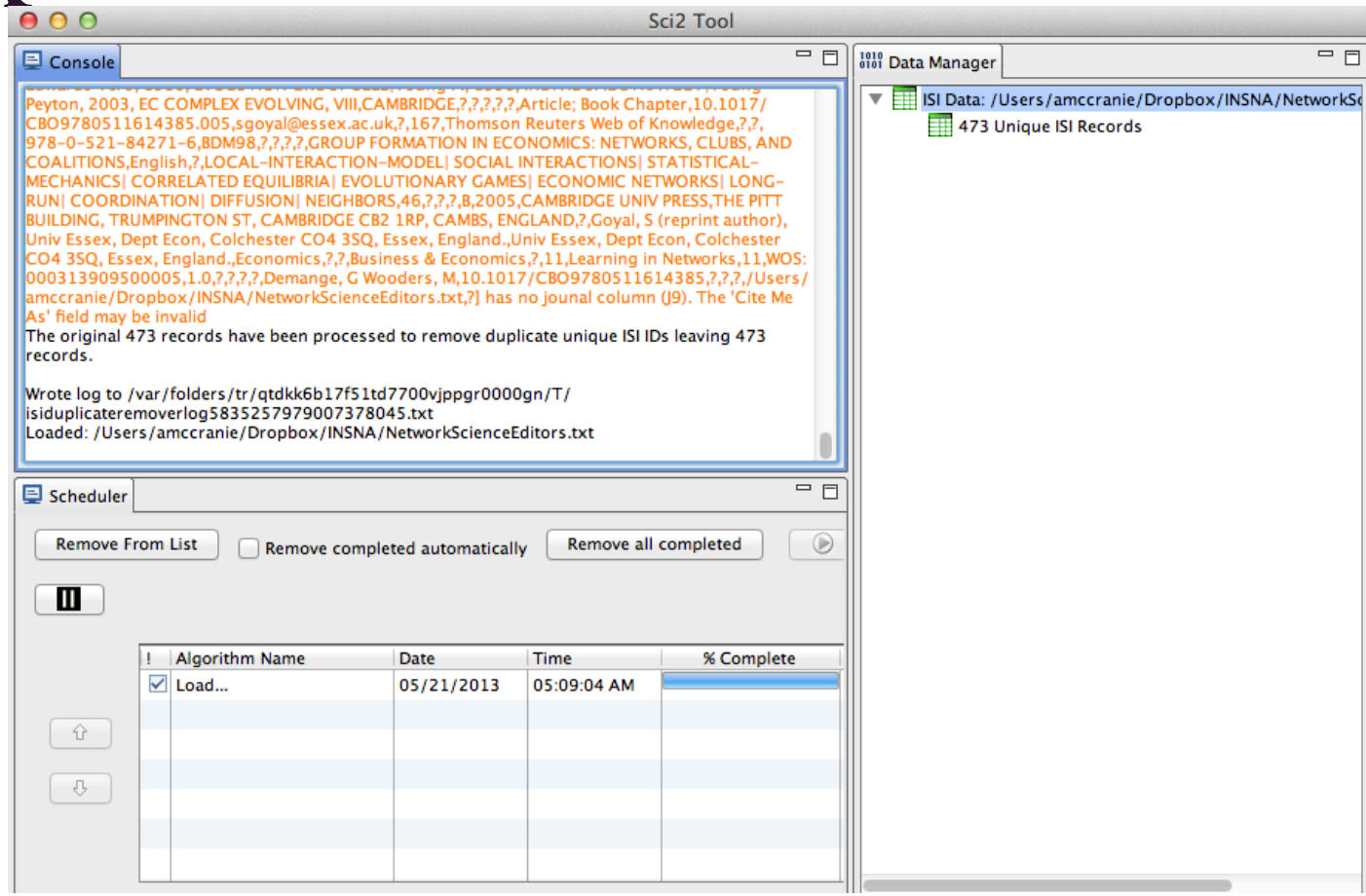
Network Science Journal Editors

- Alessandro Vespignani
- Lada Adamic
- Nosh Contractor
- Stanley Wasserman
- Thomas Valente
- Garry Robins
- Sanjeev Goyal
- Ulrik Brandes

Constructing a Co-Authorship Network

FILE>LOAD>NetworkScienceEditors.txt (in your DATA folder)

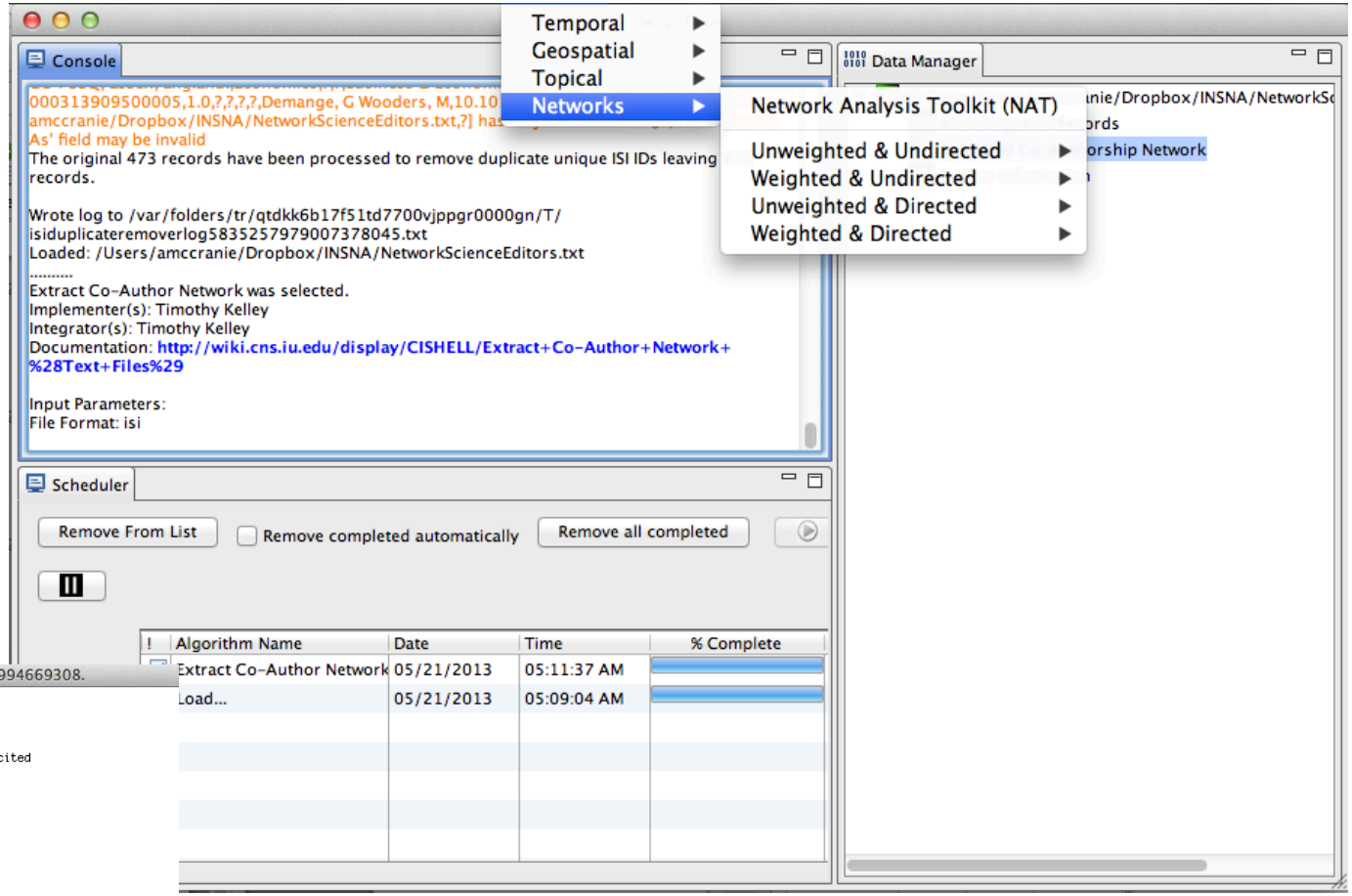
You will see some errors! These are items that are not journal articles but books and chapters. You could restart your search and exclude them.

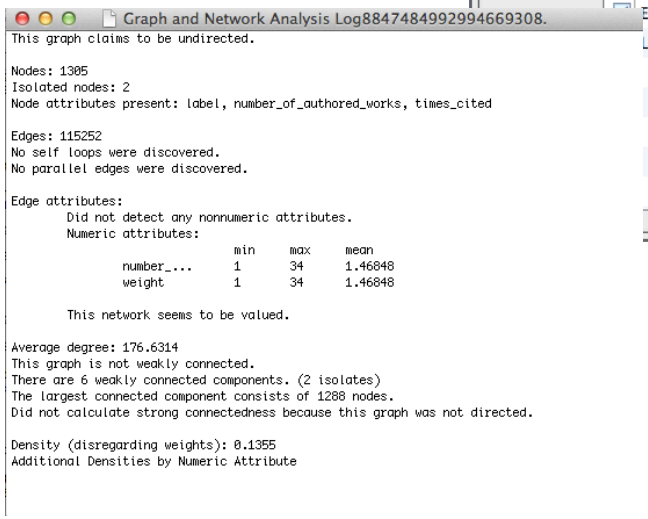

Sci2 Tool
Console
Peyton, 2003, EC COMPLEX EVOLVING, VIII,CAMBRIDGE,?,?,?,?,?,Article; Book Chapter,10.1017/CBO9780511614385.005,sgoyal@essex.ac.uk,?,167,Thomson Reuters Web of Knowledge,?,?,978-0-521-84271-6,BDM98,?,?,?,?,GROUP FORMATION IN ECONOMICS: NETWORKS, CLUBS, AND COALITIONS,English,?,LOCAL-INTERACTION-MODEL| SOCIAL INTERACTIONS| STATISTICAL-MECHANICS| CORRELATED EQUILIBRIA| EVOLUTIONARY GAMES| ECONOMIC NETWORKS| LONG-RUN| COORDINATION| DIFFUSION| NEIGHBORS,46,?,?,?,B,2005,CAMBRIDGE UNIV PRESS,THE PITT BUILDING, TRUMPINGTON ST, CAMBRIDGE CB2 1RP, CAMBS, ENGLAND,?,Goyal, S (reprint author), Univ Essex, Dept Econ, Colchester CO4 3SQ, Essex, England.,Univ Essex, Dept Econ, Colchester CO4 3SQ, Essex, England.,Economics,?,?,Business & Economics,?,11,Learning in Networks,11,WOS:000313909500005,1.0,?,?,?,?,Demange, G Wooders, M,10.1017/CBO9780511614385,?,?,?,/Users/amccranie/Dropbox/INSNA/NetworkScienceEditors.txt,?] has no jounal column (J9). The 'Cite Me As' field may be invalid
The original 473 records have been processed to remove duplicate unique ISI IDs leaving 473 records.
Wrote log to /var/folders/tr/qtdkk6b17f51td7700vjppgr0000gn/T/isiduplicateremoverlog5835257979007378045.txt
Loaded: /Users/amccranie/Dropbox/INSNA/NetworkScienceEditors.txt
Scheduler
Remove From List
Remove completed automatically
Remove all completed
! Algorithm Name Date Time % Complete
Load... 05/21/2013 05:09:04 AM
Data Manager
ISI Data: /Users/amccranie/Dropbox/INSNA/NetworkSc
473 Unique ISI Records

Constructing a Co-Authorship Network

DATA PREPARATION> EXTRACT CO-AUTHOR NETWORK.

Now, look at the network with the NAT.
Note the number of nodes.

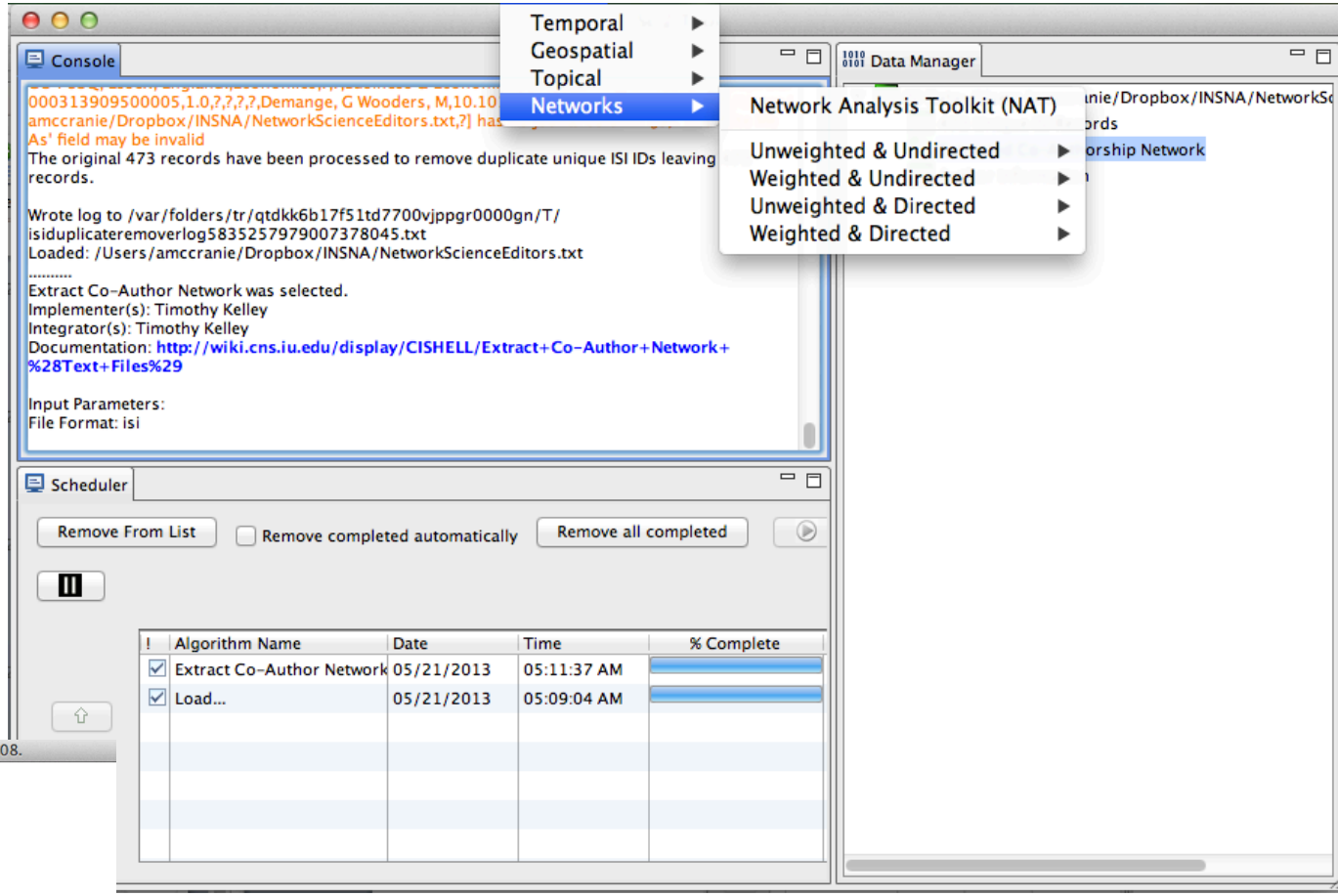


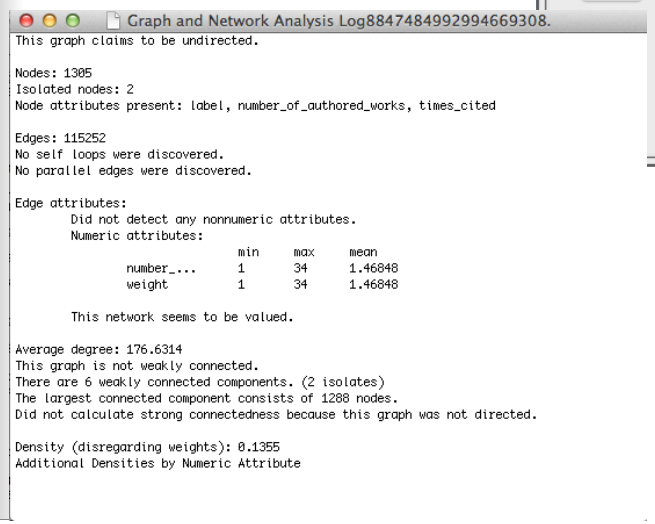

Graph and Network Analysis Log8847484992994669308.
This graph claims to be undirected.
Nodes: 1305
Isolated nodes: 2
Node attributes present: label, number_of_authored_works, times_cited
Edges: 115252
No self loops were discovered.
No parallel edges were discovered.
Edge attributes:
Did not detect any nonnumeric attributes.
Numeric attributes:
min max mean
number_... 1 34 1.46848
weight 1 34 1.46848
This network seems to be valued.
Average degree: 176.6314
This graph is not weakly connected.
There are 6 weakly connected components. (2 isolates)
The largest connected component consists of 1288 nodes.
Did not calculate strong connectedness because this graph was not directed.
Density (disregarding weights): 0.1355
Additional Densities by Numeric Attribute

Constructing a Co-Authorship Network

DATA PREPARATION> EXTRACT CO-AUTHOR NETWORK.

Now, look at the network with the NAT.
Note the number of nodes.




Graph and Network Analysis Log8847484992994669308.
This graph claims to be undirected.
Nodes: 1305
Isolated nodes: 2
Node attributes present: label, number_of_authored_works, times_cited
Edges: 115252
No self loops were discovered.
No parallel edges were discovered.
Edge attributes:
Did not detect any nonnumeric attributes.
Numeric attributes:
min max mean
number_... 1 34 1.46848
weight 1 34 1.46848
This network seems to be valued.
Average degree: 176.6314
This graph is not weakly connected.
There are 6 weakly connected components. (2 isolates)
The largest connected component consists of 1288 nodes.
Did not calculate strong connectedness because this graph was not directed.
Density (disregarding weights): 0.1355
Additional Densities by Numeric Attribute

Constructing a Co-Authorship Network

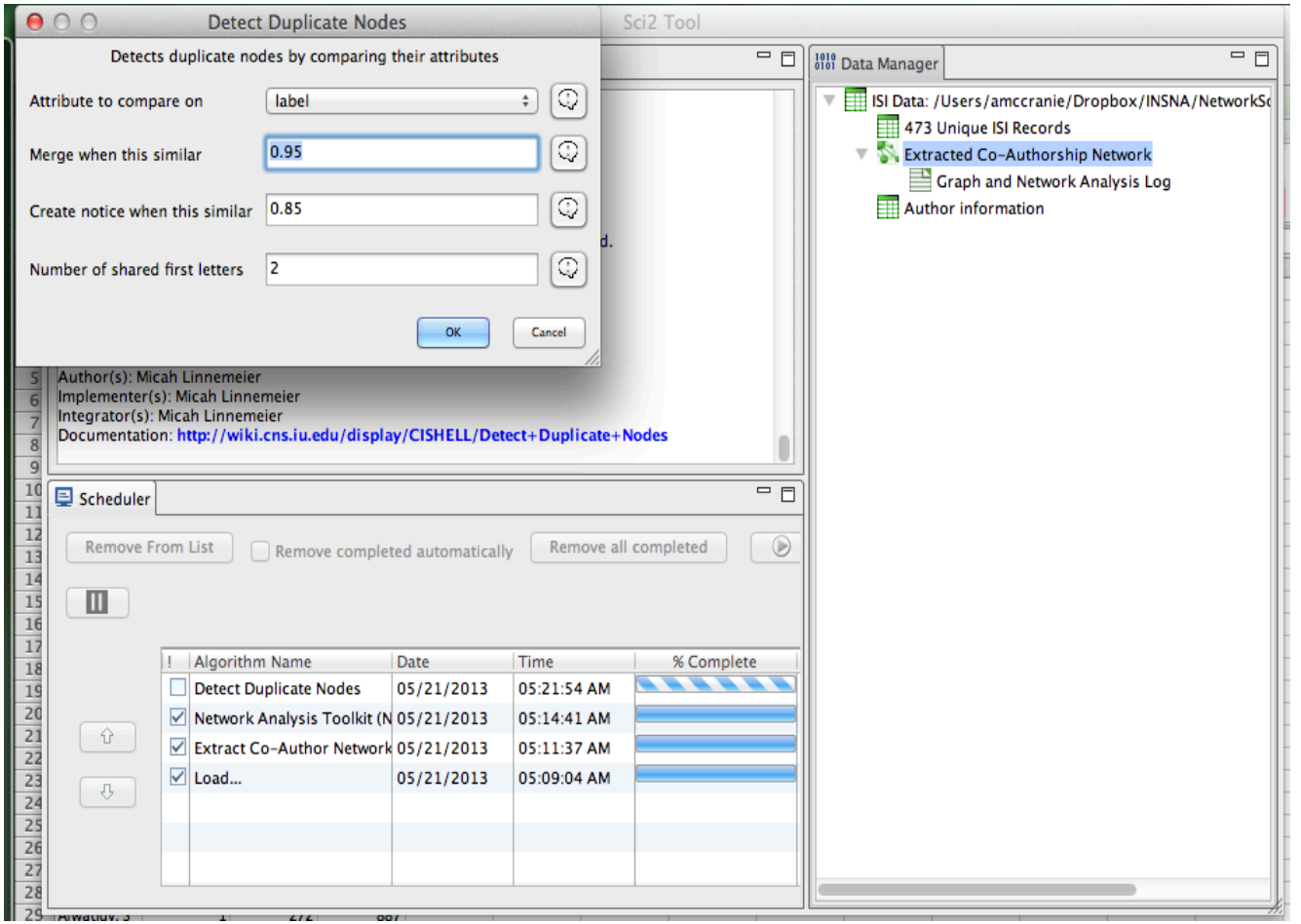
Right click on Author Information in the Data Manager and view the file.

Sort by number of articles and by label. Look for Wasserman, for instance. Notice that he is in there more than once with different name spellings. Adamic is, too. This is a problem!

So, choose the network again in the data manager and

DATA PREPARATION>DETECT DUPLICATE NODES

4	Abernethy, J	1	272	1083	*
5	Adamic, L	2	283	398	*
6	Adamic, La	9	829	473	*
7	Adar, E	2	236	605	*
8	Adeolu, A	1	272	860	*
9	Adevemi-dor	1	272	862	*

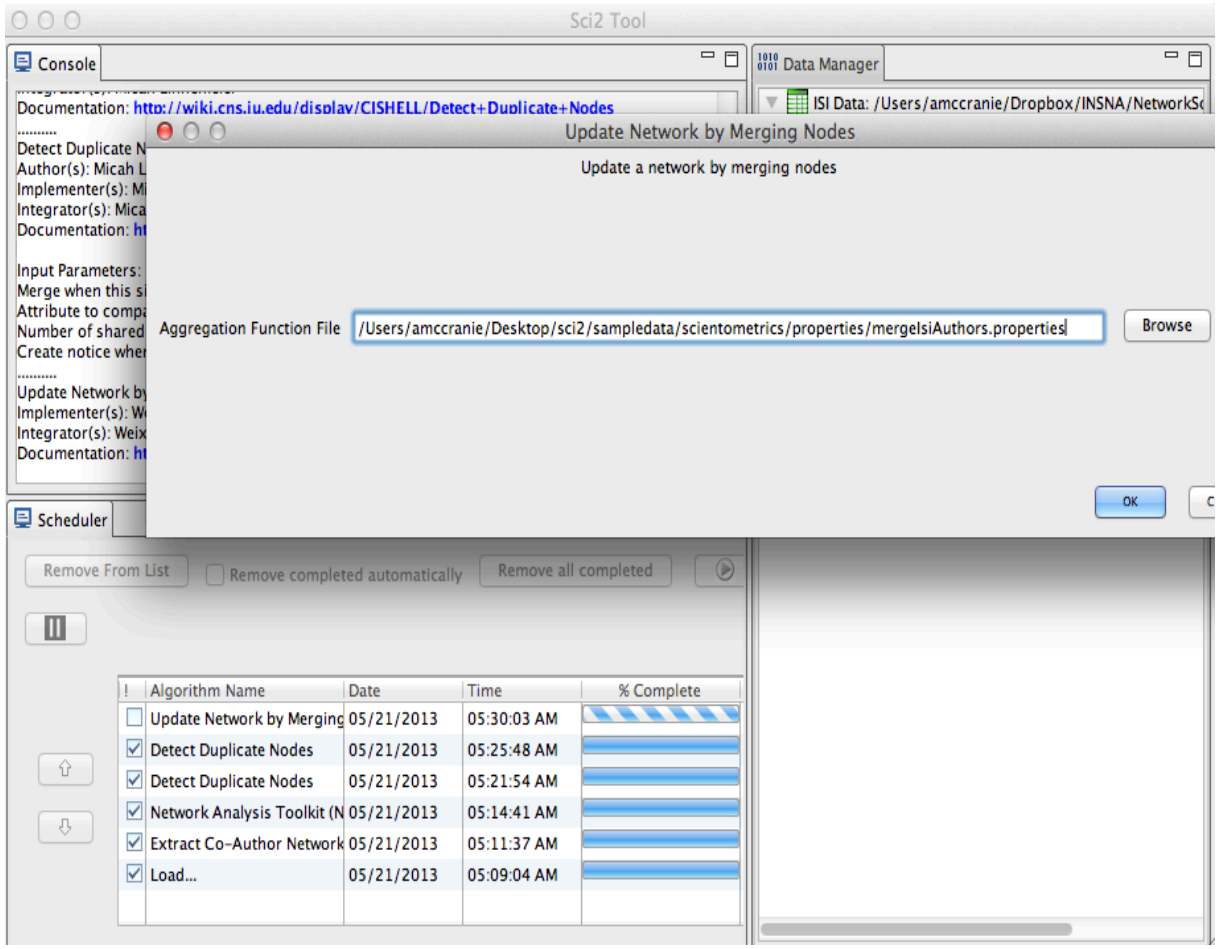


Constructing a Co-Authorship Network

Now examine the merge table. Look at the two reports. It's clearly not perfect, but it's a start. In actual analysis, you would want to be as precise as possible.

But for now, CNTRL-Click the Network and the Merge Table and go to DATA PREPARATION> UPDATE NETWORK BY MERGING NODES

For this, you need the aggregation function file shown.



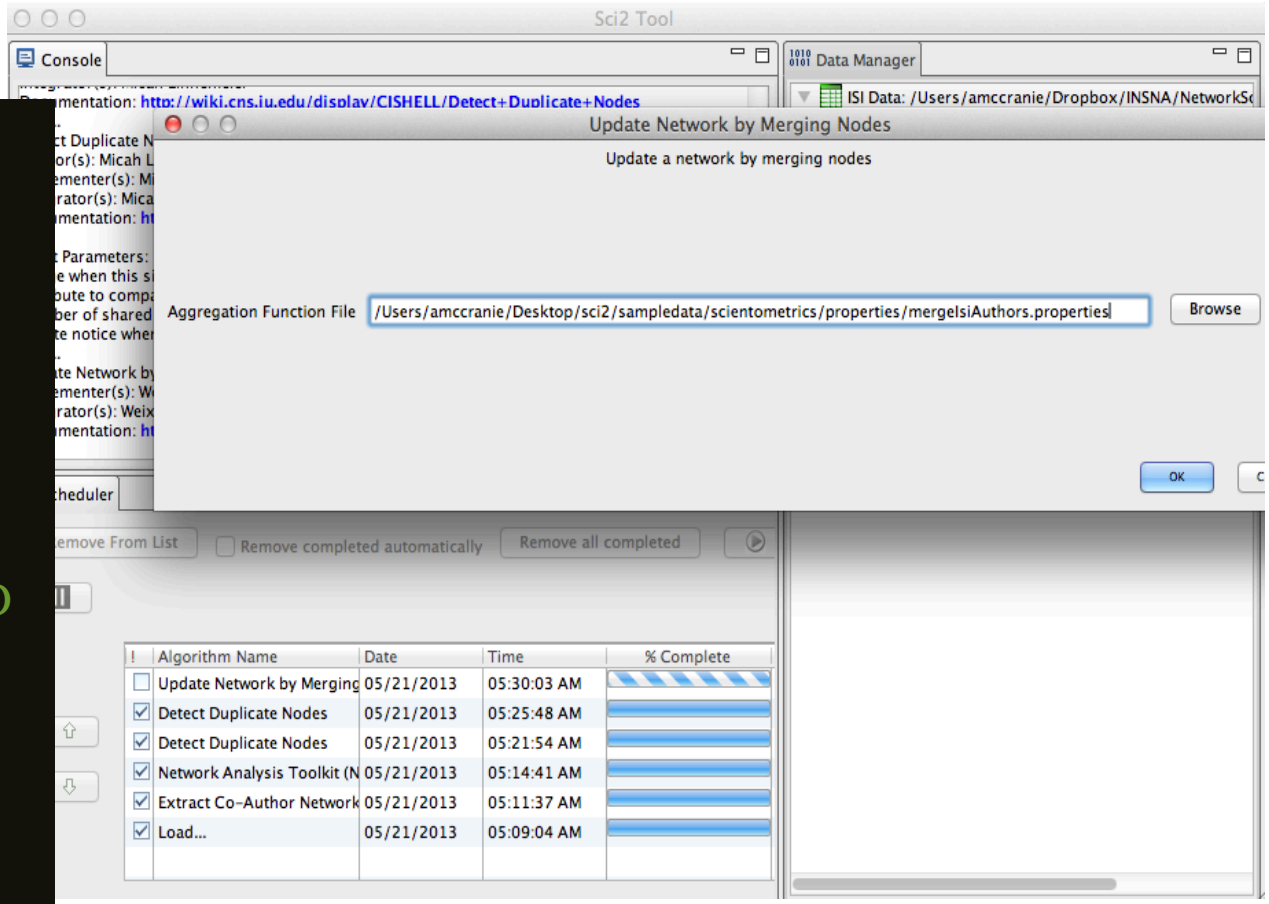
Constructing a Co-Authorship Network

Take a look at the number of nodes (NAT). Compare to the earlier number.

It should be lower than it was.

If you would like, you can separate the node and edge files with FILE> SPLIT NODE AND EDGE FILES. You can later merge them together for a new network.

This can make it easier to examine and to manipulate or add columns for nodes.

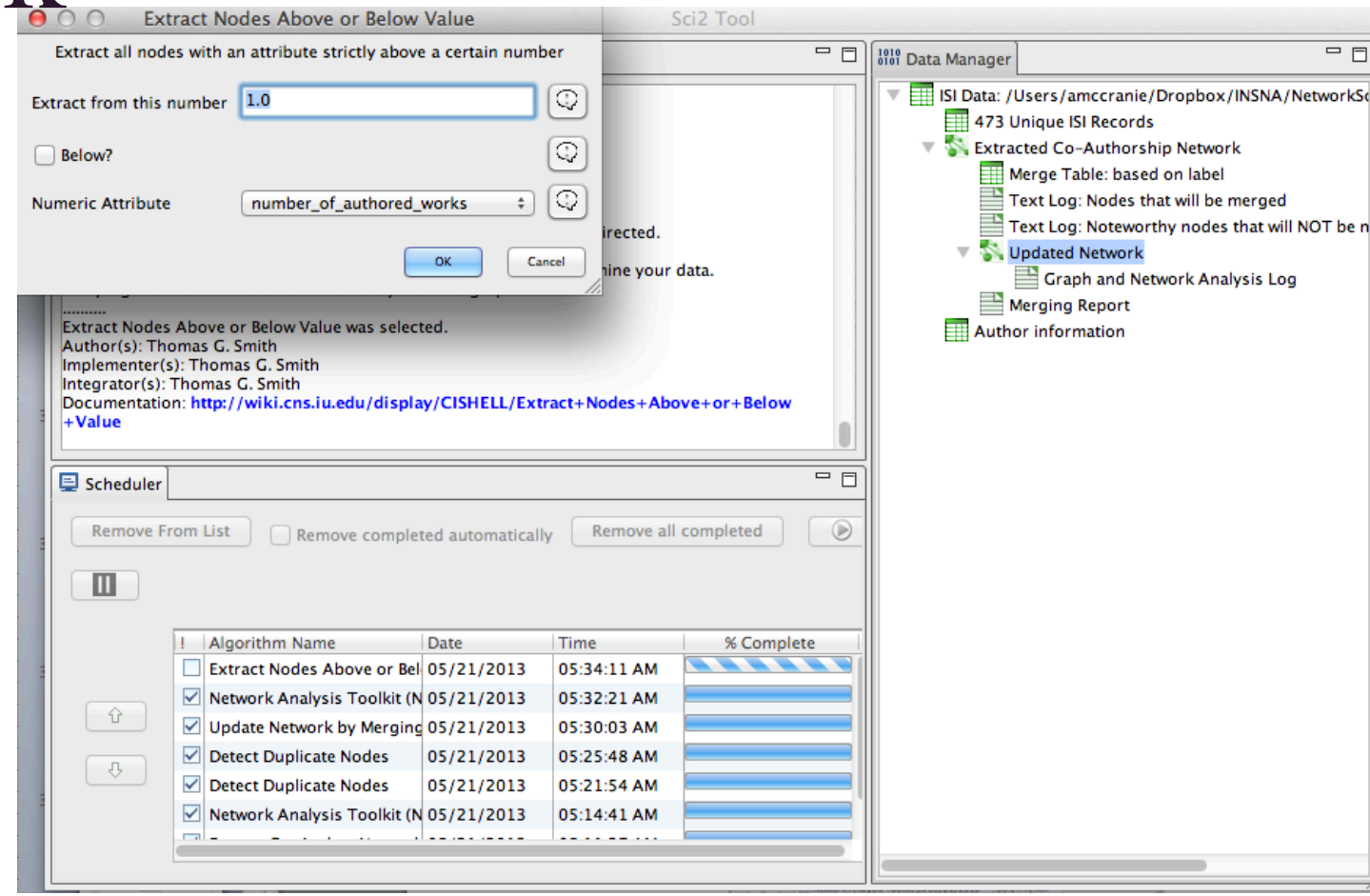


Constructing a Co-Authorship Network

Now we will remove all authors that published fewer than 3 papers in this dataset.

PREPROCESSING >NETWORKS>EXTRACT NODES ABOVE OR BELOW VALUE. (Enter 2 instead of the 3 you see.)

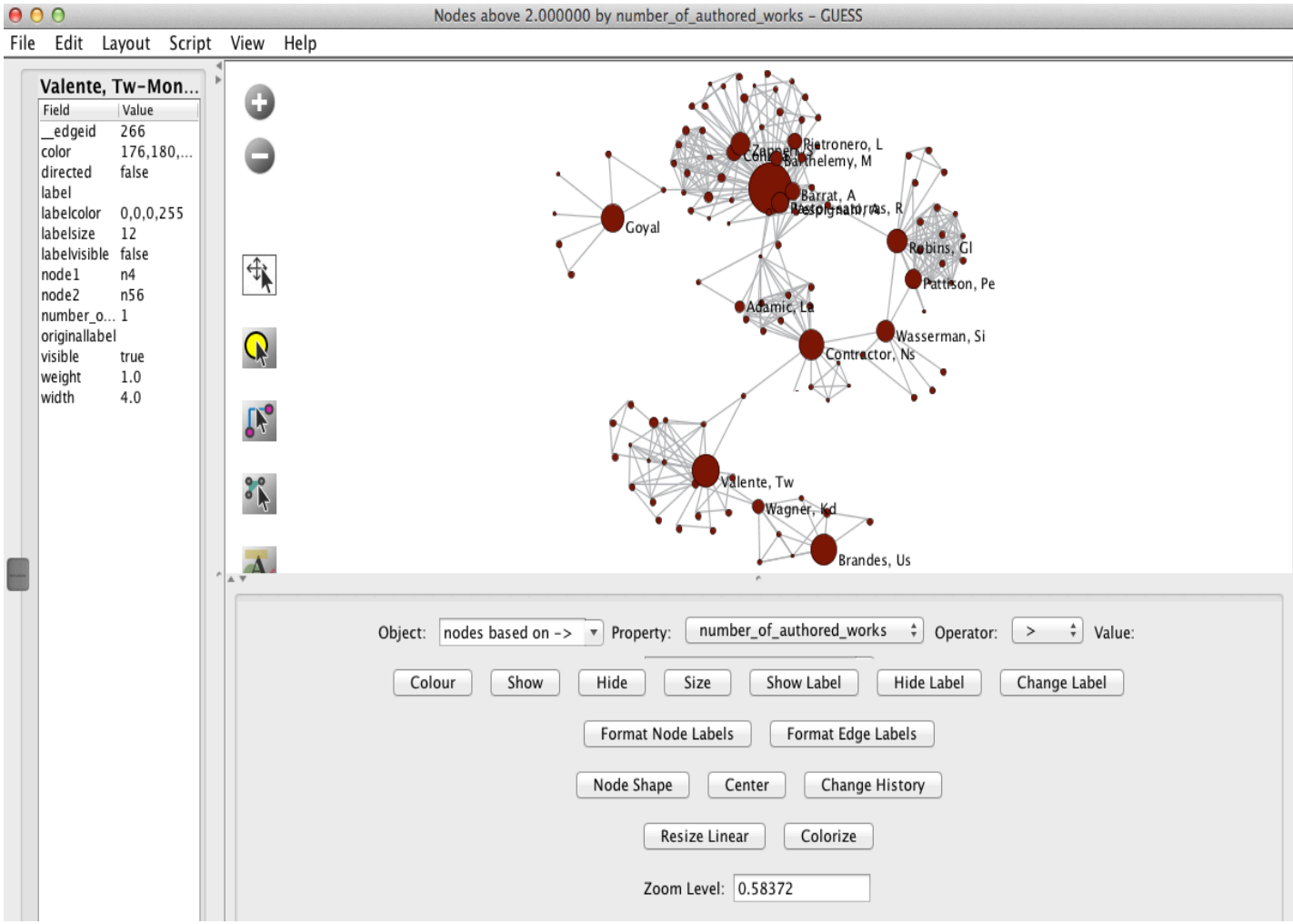
Run the NAT again and see what has changed.

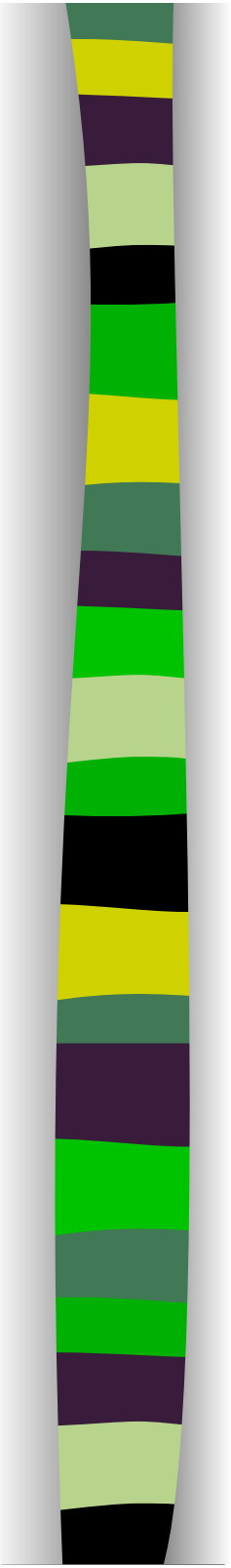


Constructing a Co-Authorship Network

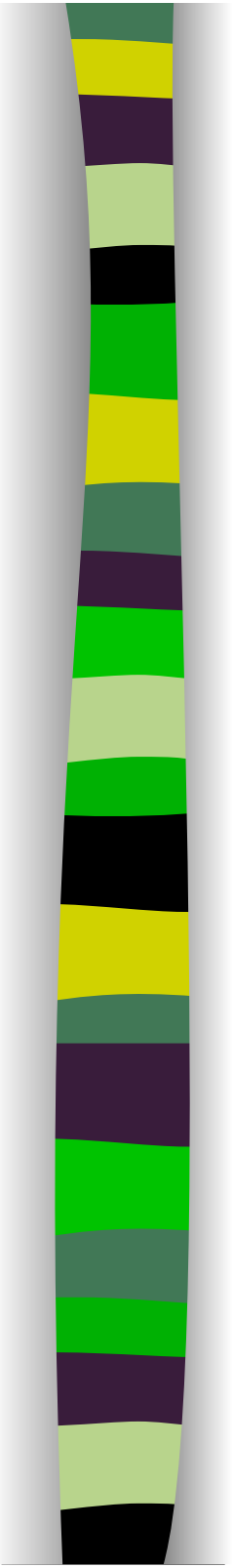
Let's visualize in GUESS.

I recommend Kamada Kawai. Resize Linear for the nodes based on the number of articles written. I recommend 1 to 50 for scale, but experiment. Show only the labels for authors of over 10 papers by choosing Nodes Based On.





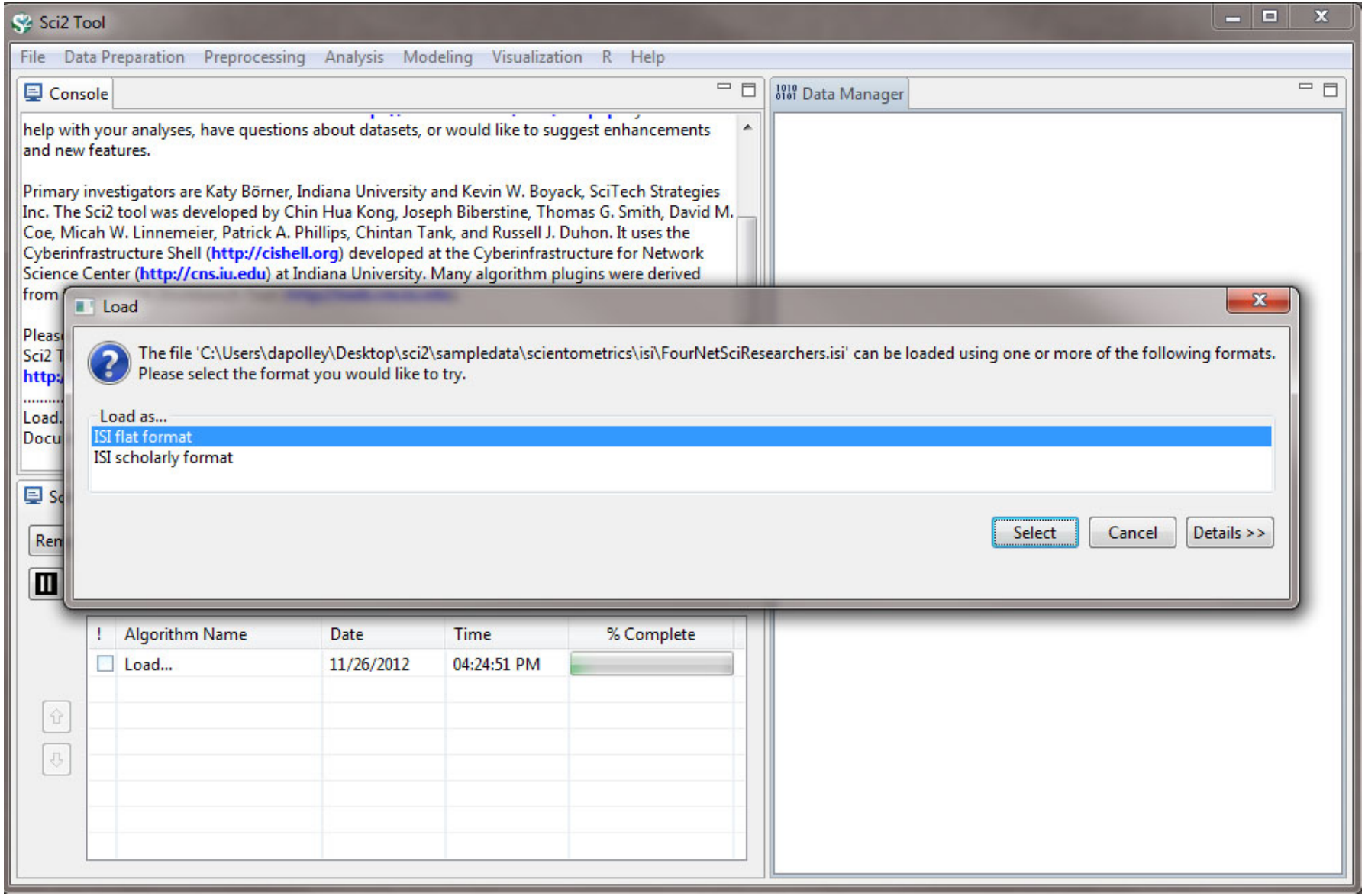
USING A MAP OF SCIENCE TO "LOCATE" RESEARCH



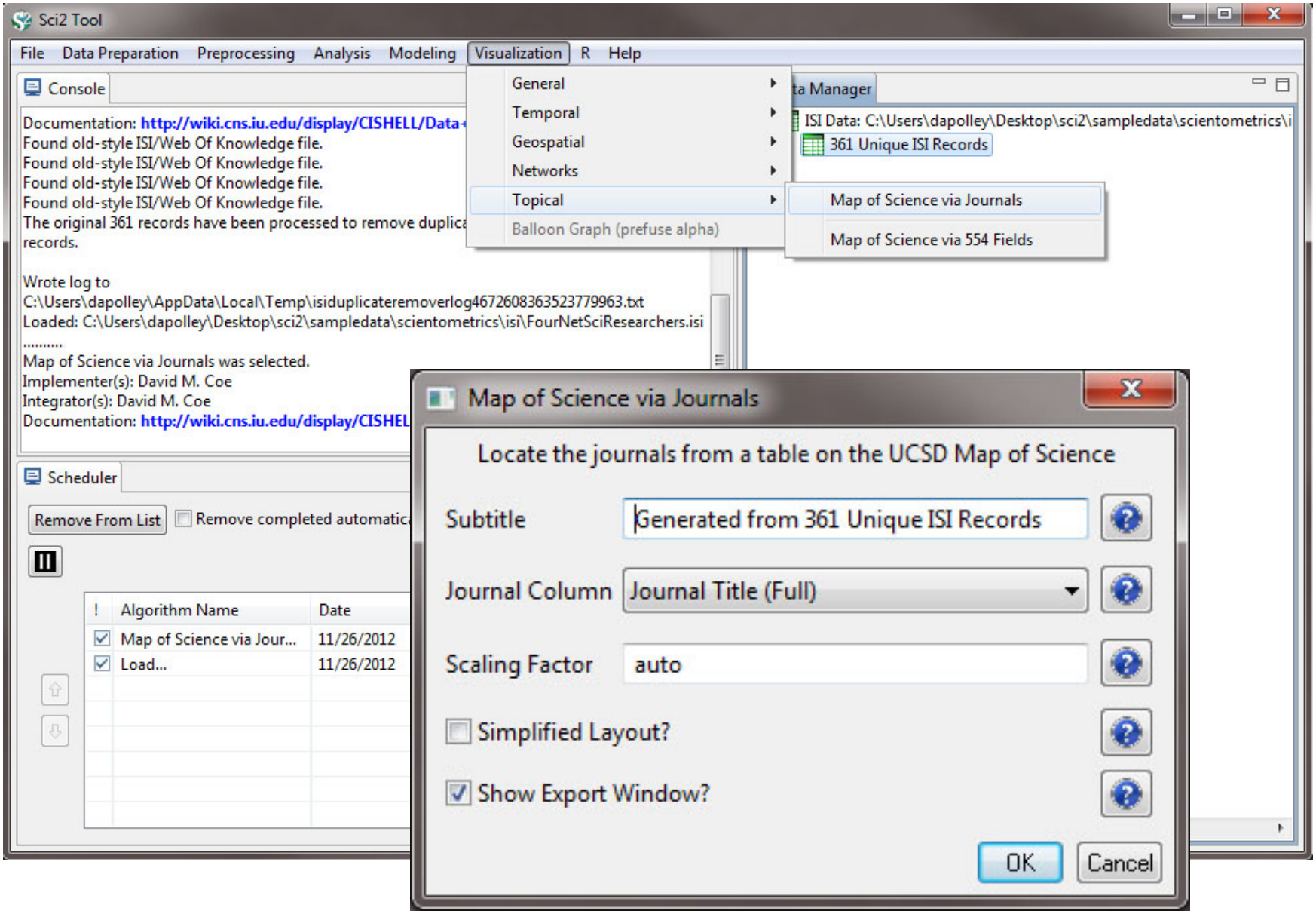
The Map of Science is a visual representation of 554 sub-disciplines within 13 disciplines of science and their relationships to one another, shown as points and lines connecting those points respectively. Over top this visualization is drawn the result of mapping a dataset's journals to the underlying sub-discipline(s) those journals contain. Mapped sub-disciplines are shown with size relative to the number matching journals and color from the discipline. For more information on maps of science, see http://mapofscience.com

As of the Sci2 v1.0 alpha release there is a plugin for Sci2 that allows users to visualize their own data overlaid on the Map of Science.

Load the FourNetSciResearchers.isi file in the ISI flat format…


Sci2 Tool
File Data Preparation Preprocessing Analysis Modeling Visualization R Help
Console
help with your analyses, have questions about datasets, or would like to suggest enhancements and new features.
Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc. The Sci2 tool was developed by Chin Hua Kong, Joseph Biberstine, Thomas G. Smith, David M. Coe, Micah W. Linnemeier, Patrick A. Phillips, Chintan Tank, and Russell J. Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University. Many algorithm plugins were derived
Data Manager
Load
The file 'C:\Users\dapolley\Desktop\sci2\sampledata\scientometrics\isi\FourNetSciResearchers.isi' can be loaded using one or more of the following formats. Please select the format you would like to try.
Load as...
ISI flat format
ISI scholarly format
Select
Cancel
Details >>
! Algorithm Name Date Time % Complete
Load... 11/26/2012 04:24:51 PM

To visualize dataset overlaid on the Map of Science run *Visualization > Topical > Map of Science via Journals*

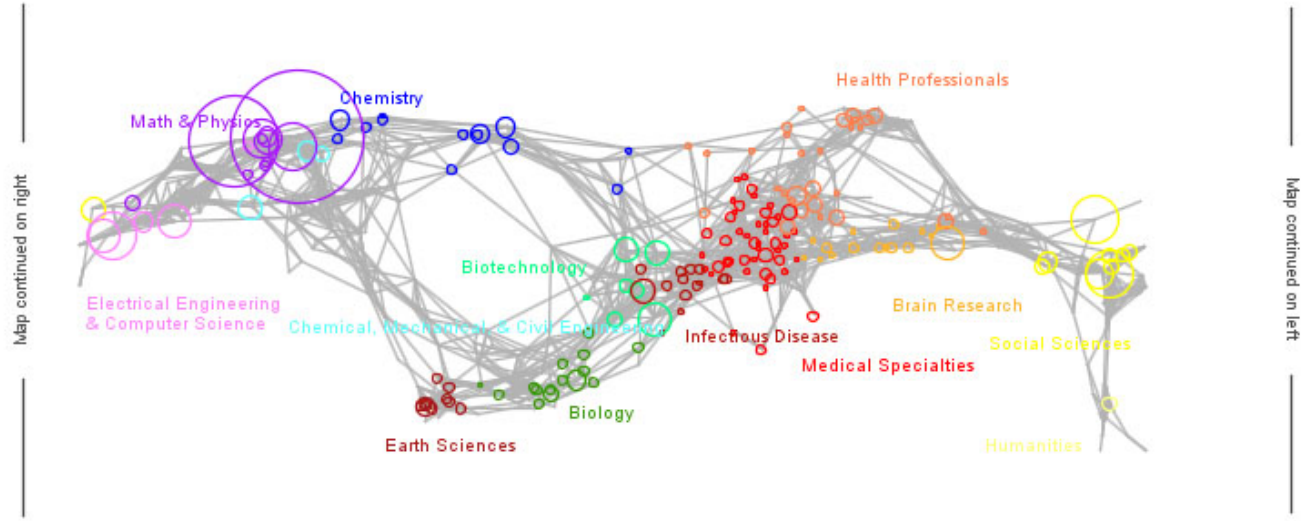


Topical Visualization

File

Topical Visualization

Generated from 361 Unique ISI Records
90 out of 112 records were mapped to 182 subdisciplines and 13 disciplines.
November 26, 2012 | 04:37 PM EST

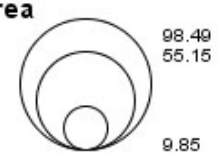


© 2008 The Regents of the University of California and SciTech Strategies.
Map updated by SciTech Strategies, OST, and CNS in 2011.

Legend

Circle area: Fractional record count
Unclassified = 22
Minimum = 0
Maximum = 98
Scaling factor = 0.5076673
Color: Discipline
See end of PDF for color legend.

Area

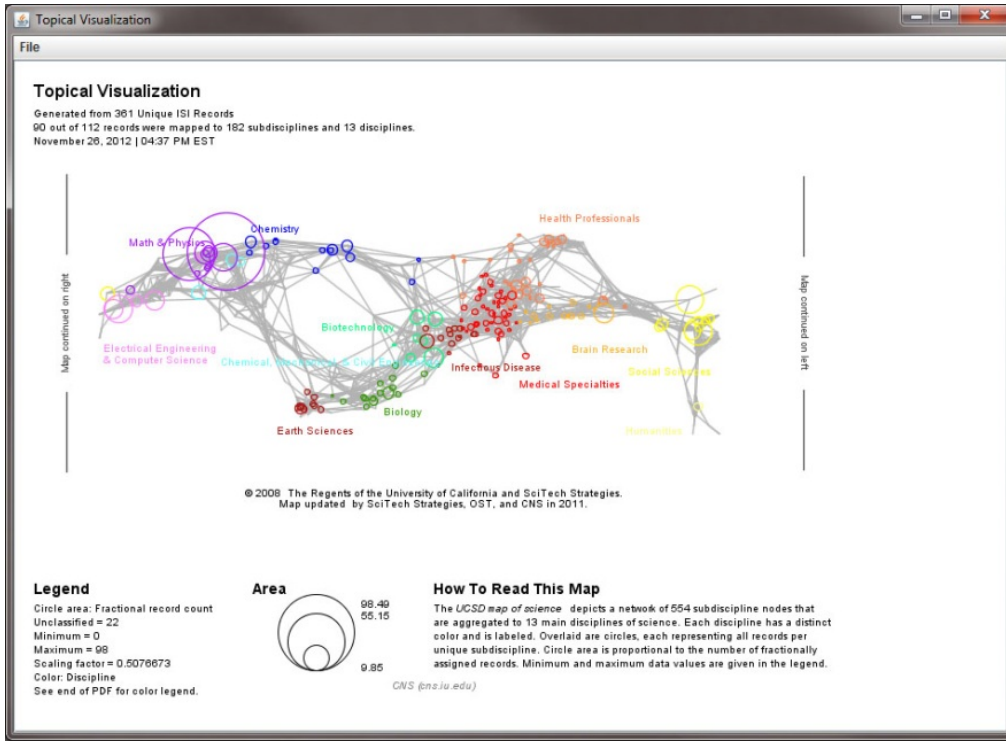


How To Read This Map

The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

CNS (cns.iu.edu)

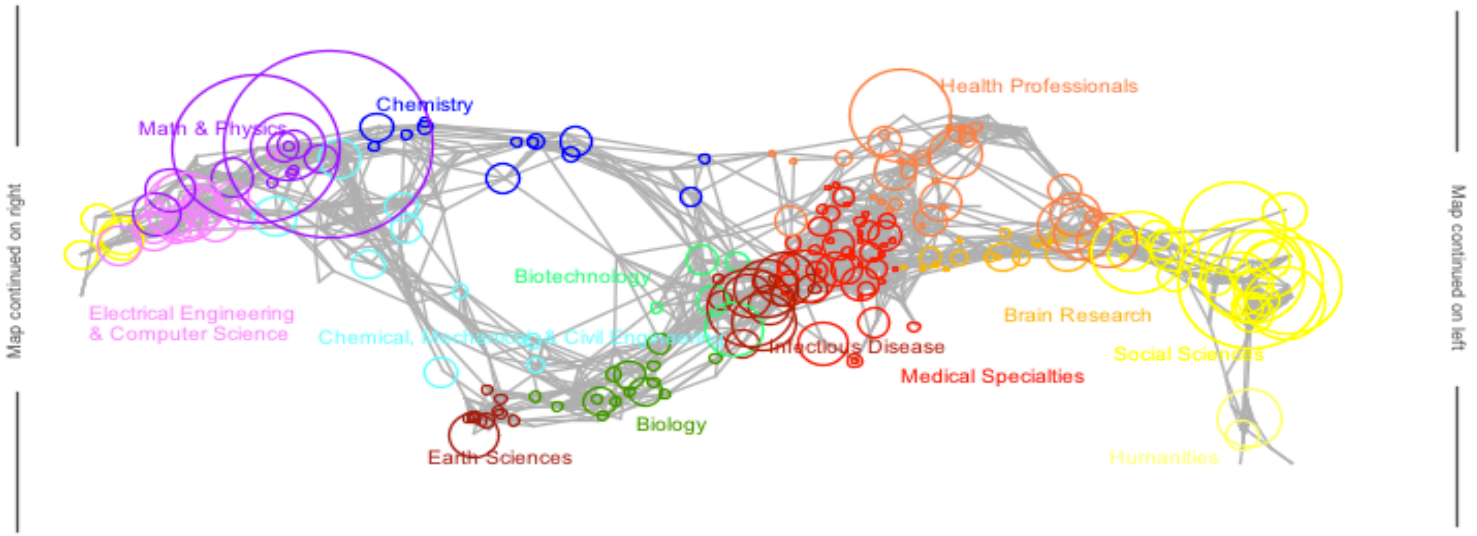
The journals titles are used to determine which records fit into what sub-discipline. You can view the journal titles found and those not found from the data manager of Sci2. A single journal can belong to more than one sub-discipline and thus so can the record associated with that journal. So the circle sizes are proportional to the number of fractionally assigned records.


Topical Visualization
File
Topical Visualization
Generated from 361 Unique ISI Records
90 out of 112 records were mapped to 182 subdisciplines and 13 disciplines.
November 26, 2012 | 04:37 PM EST
Map continued on right
Map continued on left
Math & Physics
Chemistry
Health Professionals
Biotechnology
Electrical Engineering & Computer Science
Chemical Engineering & Chemistry
Infectious Disease
Brain Research
Medical Specialties
Social Sciences
Biology
Earth Sciences
Humanities
© 2008 The Regents of the University of California and SciTech Strategies.
Map updated by SciTech Strategies, OST, and CNS in 2011.
Legend
Circle area: Fractional record count
Unclassified = 22
Minimum = 0
Maximum = 98
Scaling factor = 0.5076673
Color: Discipline
See end of PDF for color legend.
Area
98.49
55.15
9.85
How To Read This Map
The UCSD map of science depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.
CNS (cns.iu.edu)

Now, consider this with the NetworkScienceEditors.txt file. See the spread of the area coverage? That's what they were going for. Note where the largest circle is reflects the large number of contributions of Vespigani.

Topical Visualization

Generated from 473 Unique ISI Records.2
193 out of 227 publications were mapped to 222 subdisciplines and 13 disciplines.
May 21, 2013 | 06:01 AM CEST

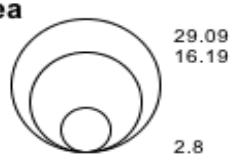


© 2008 The Regents of the University of California and SciTech Strategies.
Map updated by SciTech Strategies, OST, and CNS in 2011.

Legend

Circle area: Fractional Journal Count
Unclassified = 34
Minimum = 0
Maximum = 60
Color: Discipline
See end of PDF for color legend.

Area



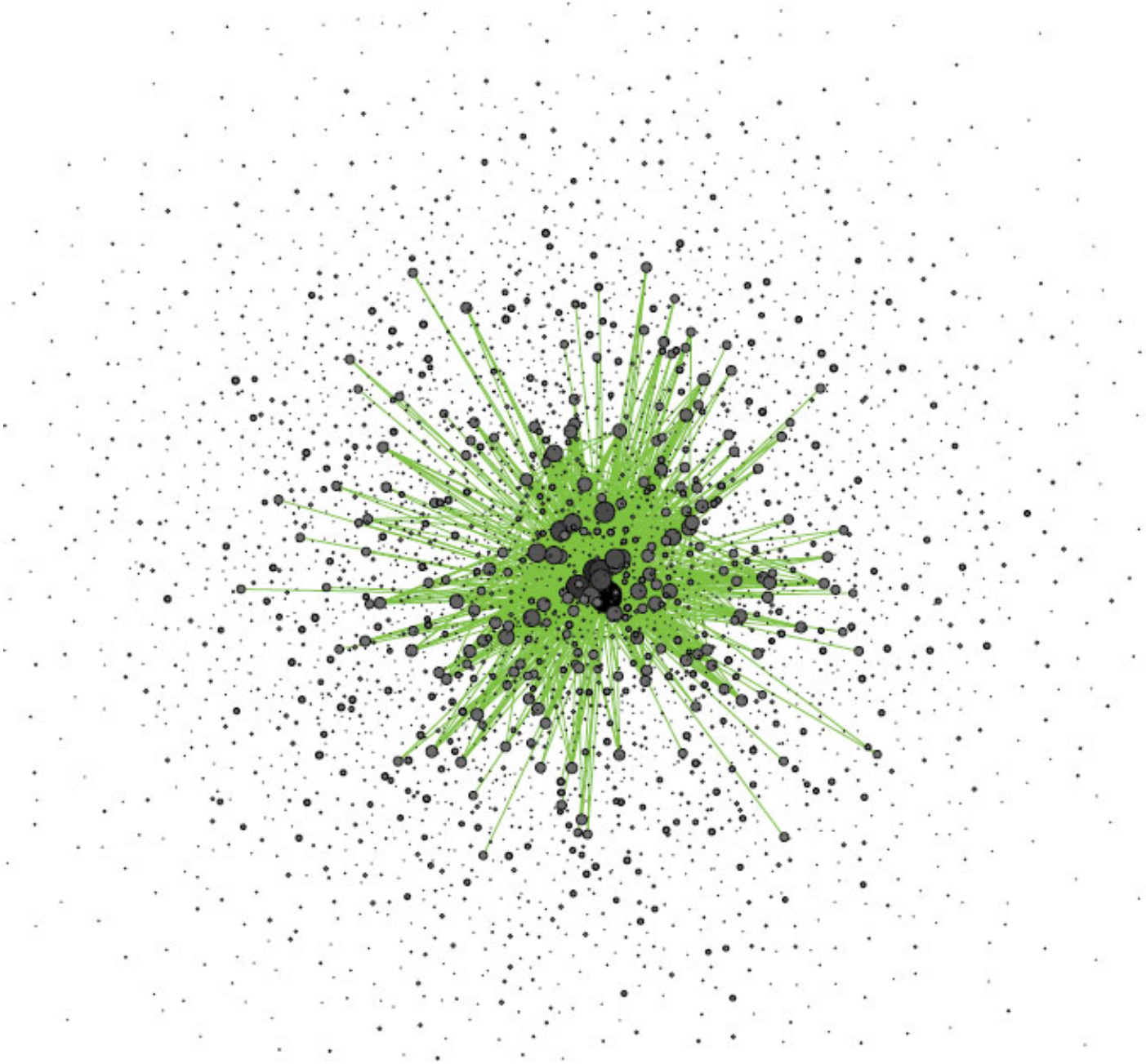
How To Read This Map

The *UCSD map of science* depicts a network of 554 subdiscipline nodes that are aggregated to 13 main disciplines of science. Each discipline has a distinct color and is labeled. Overlaid are circles, each representing all records per unique subdiscipline. Circle area is proportional to the number of fractionally assigned records. Minimum and maximum data values are given in the legend.

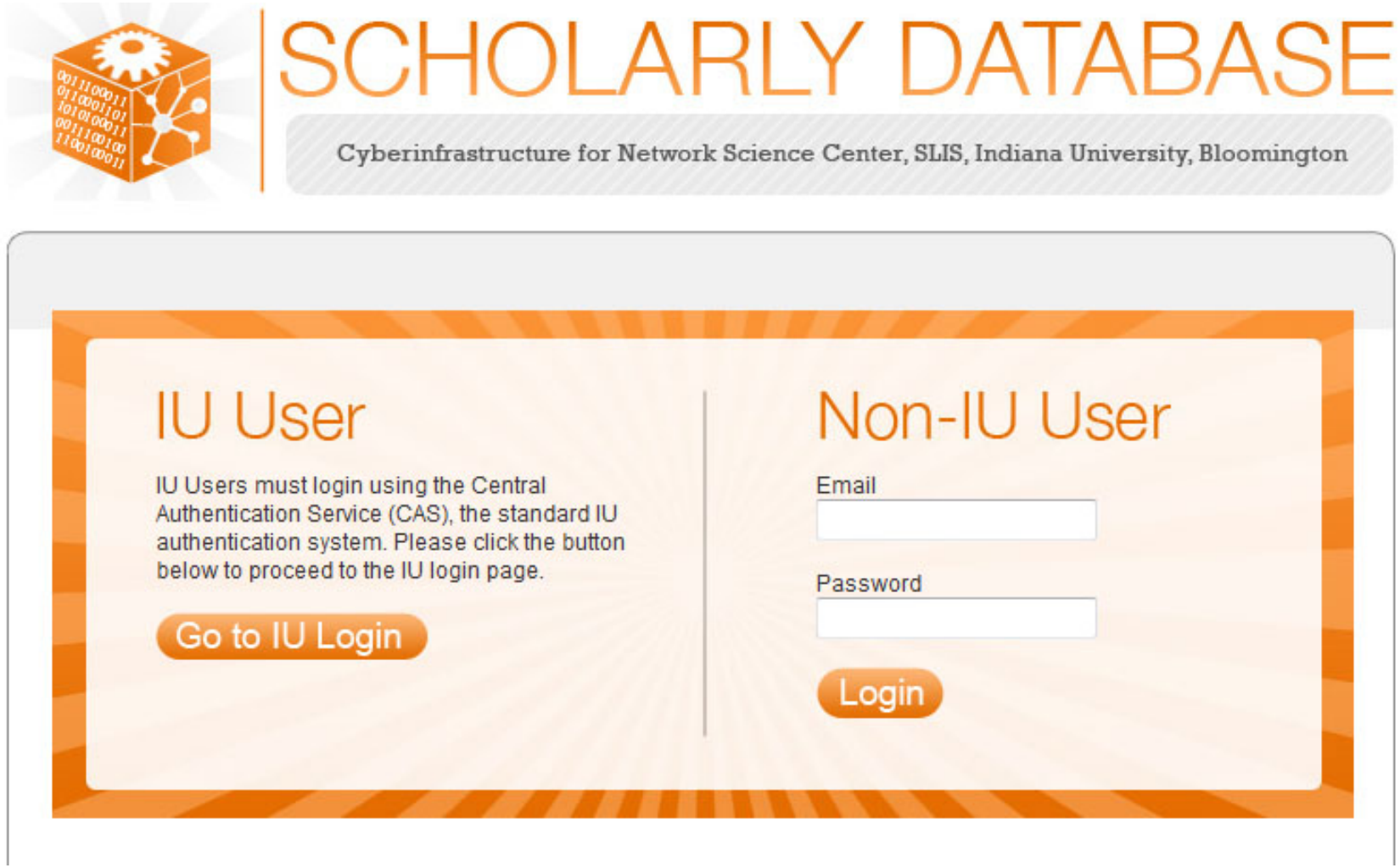
CNS (cns.iu.edu)

USING THE SCHOLARLY DATABASE ON A TOPIC AREA

Word Co-Occurrence Network from the abstracts of articles from MEDLINE with the keyword “mesothelioma” in the title…

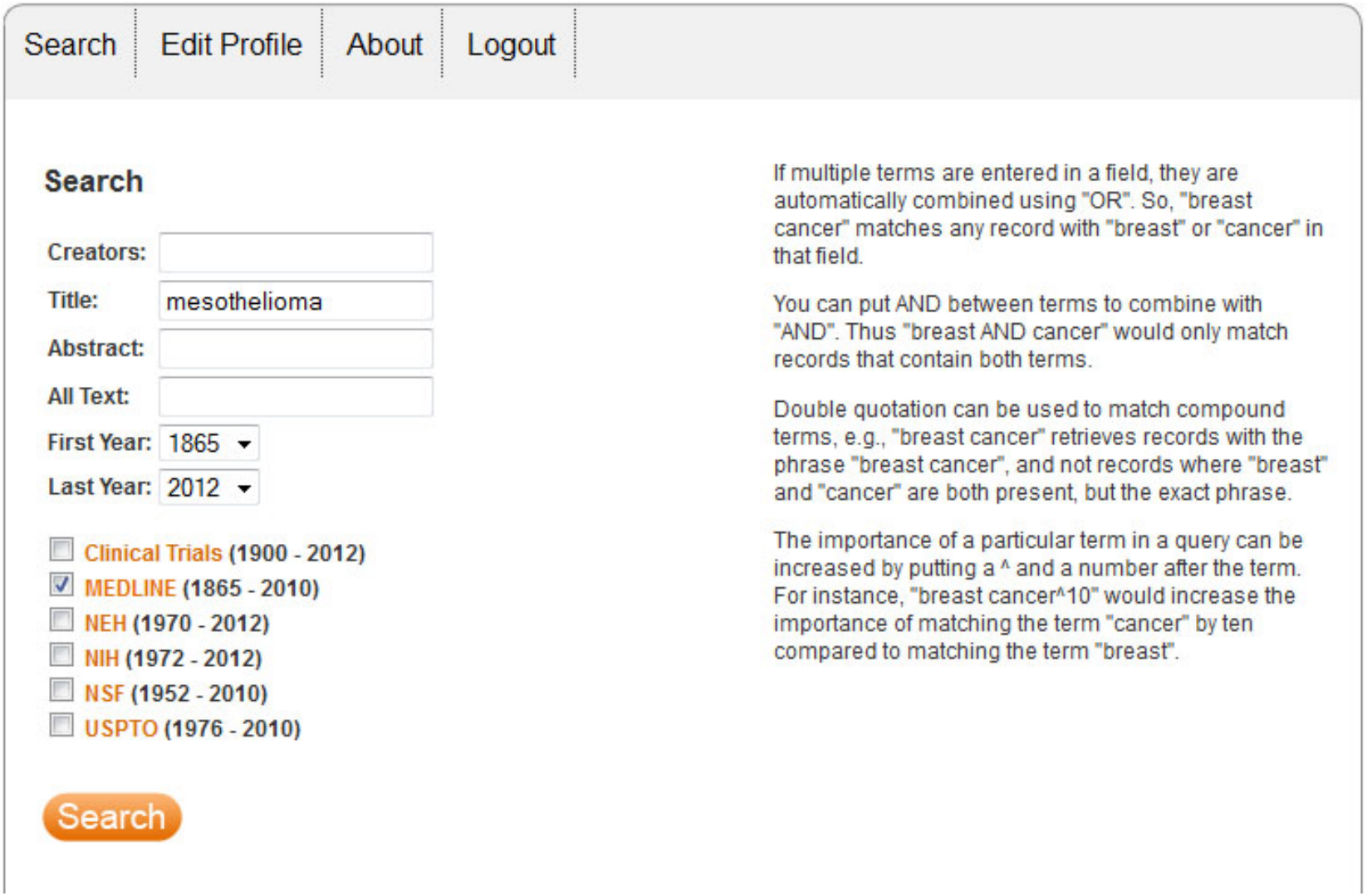


If you have registered for the Scholarly Database then go to http://sdb.cns.iu.edu and login…


SCHOLARLY DATABASE
Cyberinfrastructure for Network Science Center, SLIS, Indiana University, Bloomington
IU User
IU Users must login using the Central Authentication Service (CAS), the standard IU authentication system. Please click the button below to proceed to the IU login page.
Go to IU Login
Non-IU User
Email
Password
Login

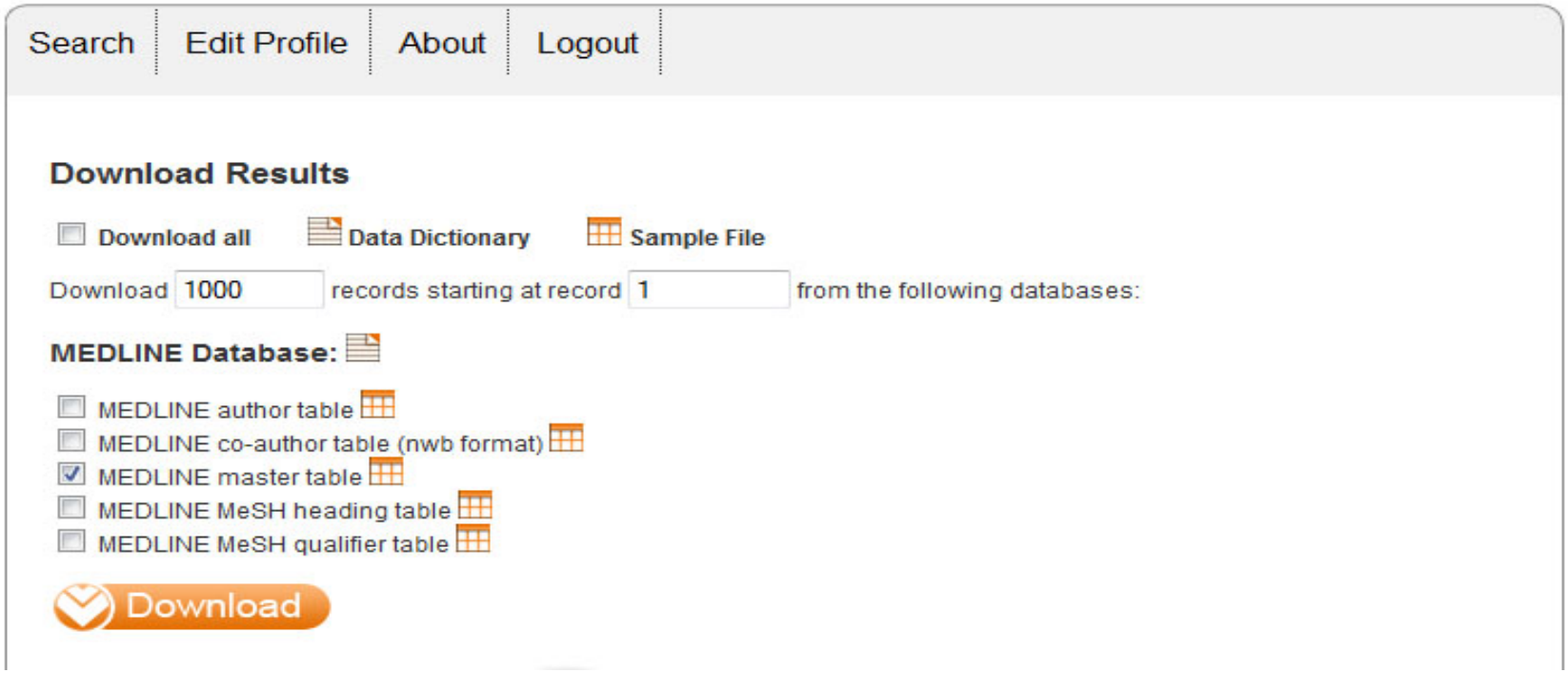
If you do not have an account type in nwb@indiana.edu and nwb for the password

Do a keyword search in Title for “mesothelioma” and check MEDLINE…


Search
Edit Profile
About
Logout
Search
Creators:
Title:
mesothelioma
Abstract:
All Text:
First Year:
1865
Last Year:
2012
Clinical Trials (1900 - 2012)
MEDLINE (1865 - 2010)
NEH (1970 - 2012)
NIH (1972 - 2012)
NSF (1952 - 2010)
USPTO (1976 - 2010)
Search
If multiple terms are entered in a field, they are automatically combined using "OR". So, "breast cancer" matches any record with "breast" or "cancer" in that field.
You can put AND between terms to combine with "AND". Thus "breast AND cancer" would only match records that contain both terms.
Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase "breast cancer", and not records where "breast" and "cancer" are both present, but the exact phrase.
The importance of a particular term in a query can be increased by putting a ^ and a number after the term. For instance, "breast cancer^10" would increase the importance of matching the term "cancer" by ten compared to matching the term "breast".

Extracting Word Co-Occurrence Network from SDB Data

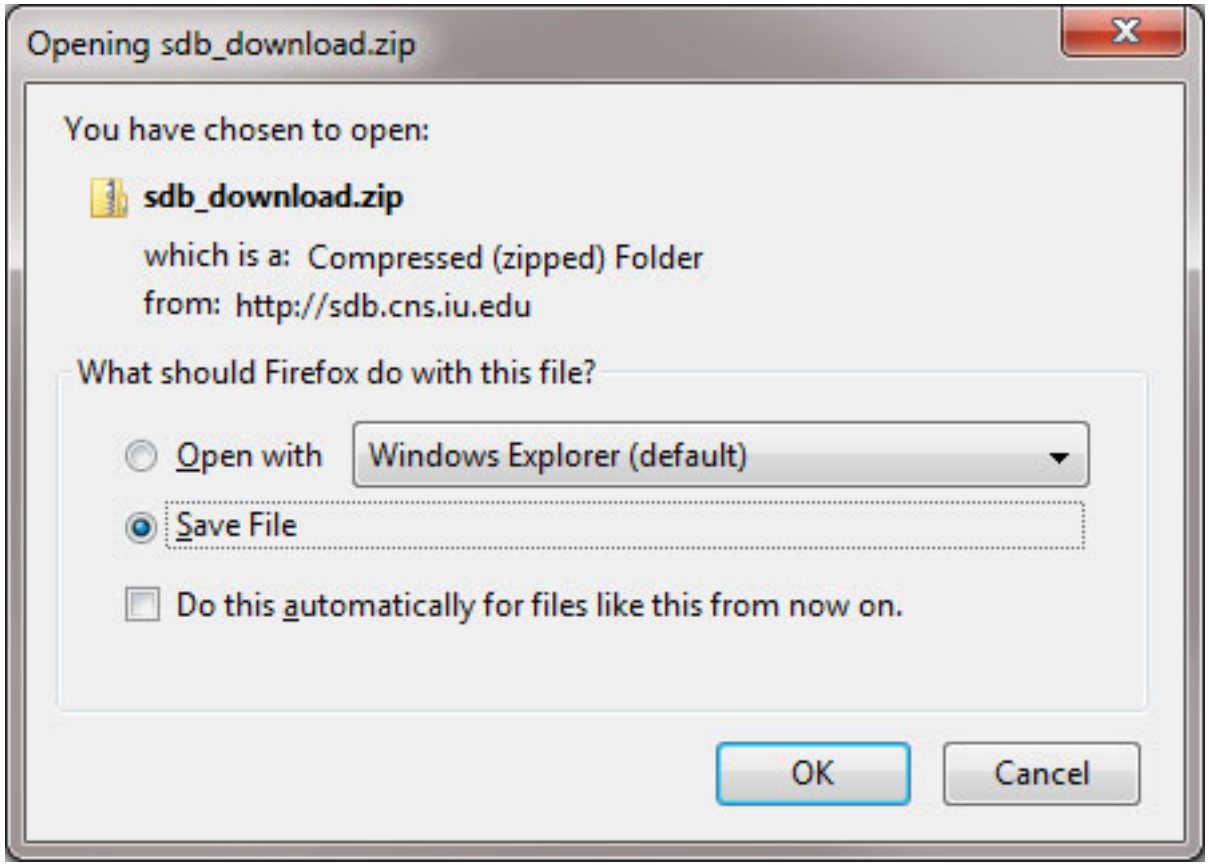
We will only download the first 1000 results to minimize the runtime for the algorithms used in this workflow. Make sure to check MEDLINE master table since that will have all of the bibliographic data we need for this analysis.



Your download limit will initially capped at 2000 records at a time. To increase this limit, please email cns-sdb-dev-l@iulist.indiana.edu

Extracting Word Co-Occurrence Network from SDB Data

Save the file somewhere on your computer for use later in this tutorial… (in this case, if you can't sign on to SDB, then you will find a copy in the DATA folder.)

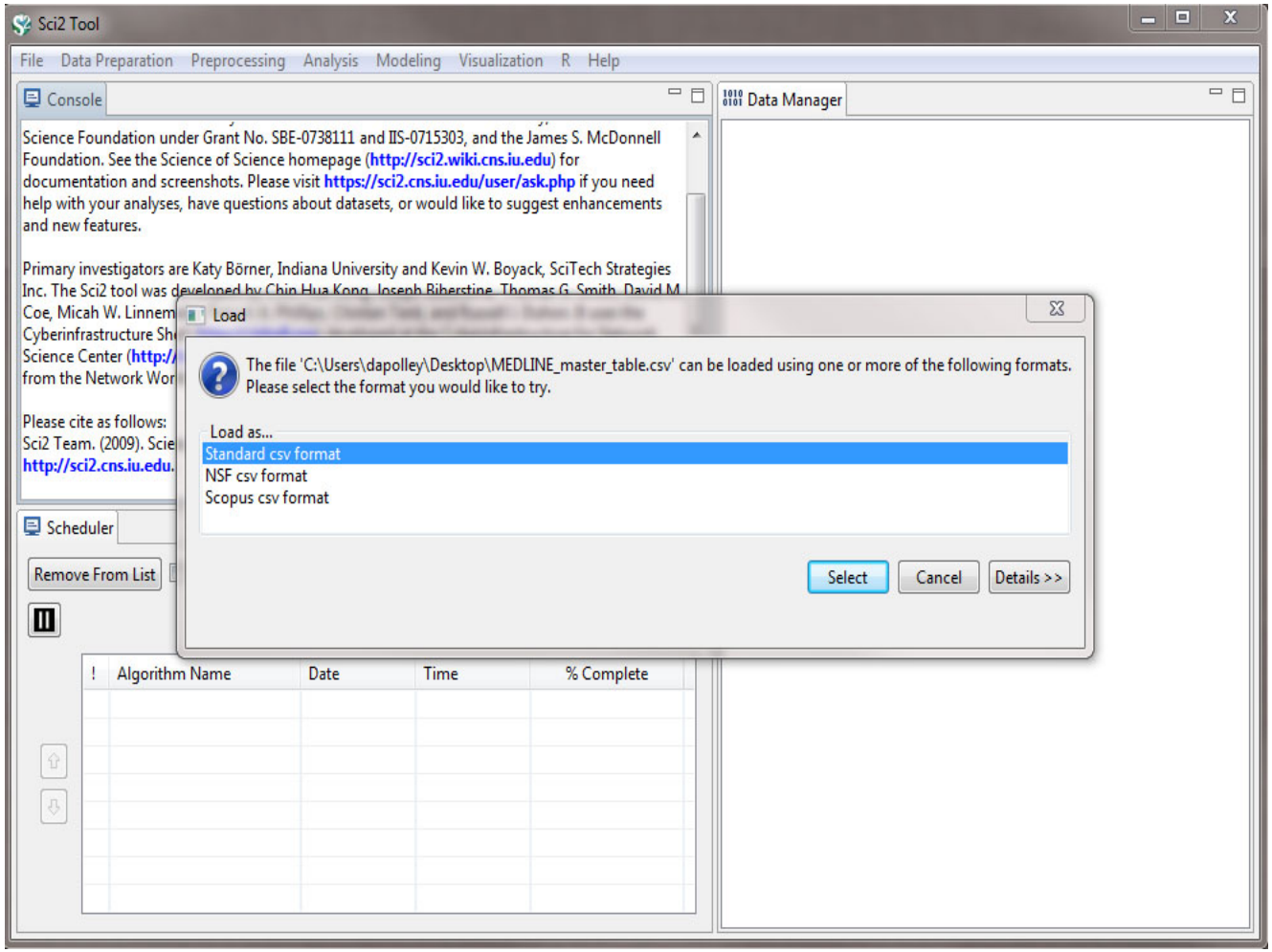


Extracting Word Co-Occurrence Network from SDB Data

- The topic similarity of basic and aggregate units of science can be calculated via an analysis of the co-occurrence of words in associated texts. Units that share more words in common are assumed to have higher topical overlap and are connected via linkages and/or placed in closer proximity.
- Extract Word Co-Occurrence Network creates a weighted network where each node is a word and edges connect words to each other, where the strength of an edge represents how often two words occur in the same body of text together.
- This algorithm is a shortcut for extracting a directed network using Extract Directed Network, and then performing bibliographic coupling using Extract Reference Co-Occurrence (Bibliographic Coupling) Network.

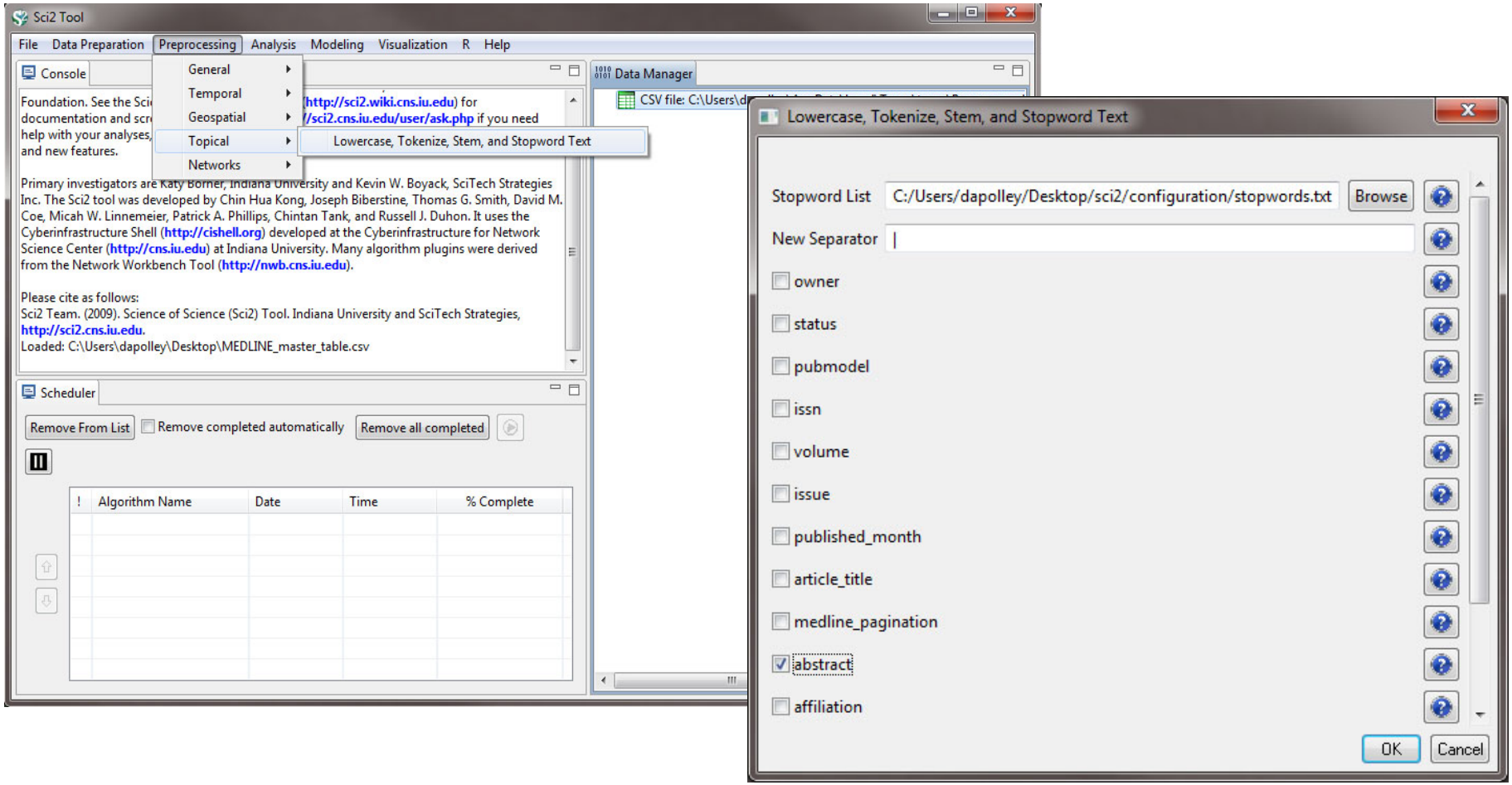
Extracting Word Co-Occurrence Network from SDB Data

Open Sci2 and load the MEDLINE_master_table.csv file as a Standard CSV file…


Sci2 Tool
File Data Preparation Preprocessing Analysis Modeling Visualization R Help
Console
Science Foundation under Grant No. SBE-0738111 and IIS-0715303, and the James S. McDonnell Foundation. See the Science of Science homepage (http://sci2.wiki.cns.iu.edu) for documentation and screenshots. Please visit https://sci2.cns.iu.edu/user/ask.php if you need help with your analyses, have questions about datasets, or would like to suggest enhancements and new features.
Primary investigators are Katy Börner, Indiana University and Kevin W. Boyack, SciTech Strategies Inc.
Please cite as follows:
Data Manager
Load
The file 'C:\Users\dapolley\Desktop\MEDLINE_master_table.csv' can be loaded using one or more of the following formats. Please select the format you would like to try.
Load as...
Standard csv format
NSF csv format
Scopus csv format
Select
Cancel
Details >>
Scheduler
Remove From List
! Algorithm Name Date Time % Complete

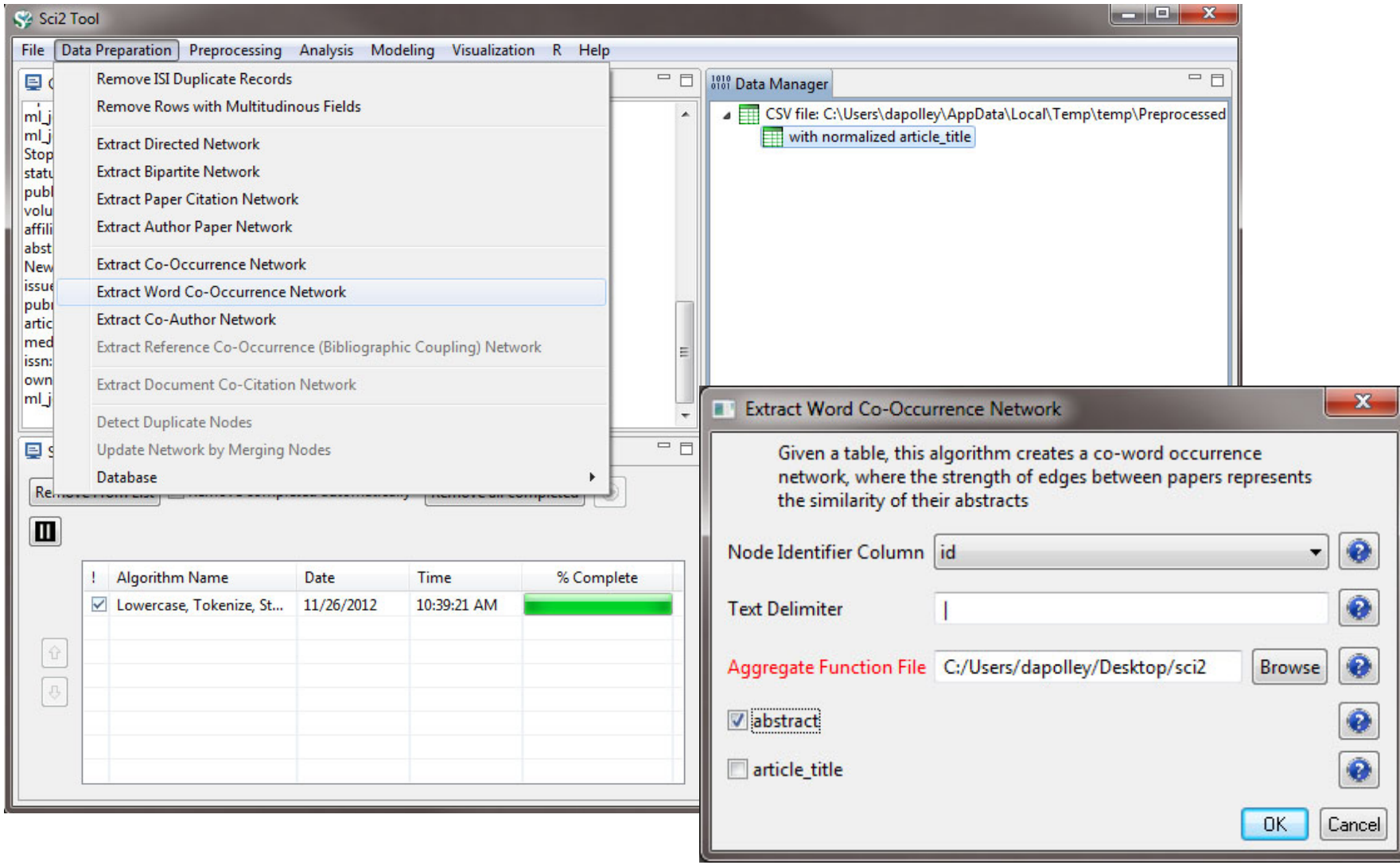
Extracting Word Co-Occurrence Network from SDB Data

Normalize the titles by running *Preprocessing > Topical > Lowercase, Tokenize, Stem, and Stopword Text* and select “Abstract”…



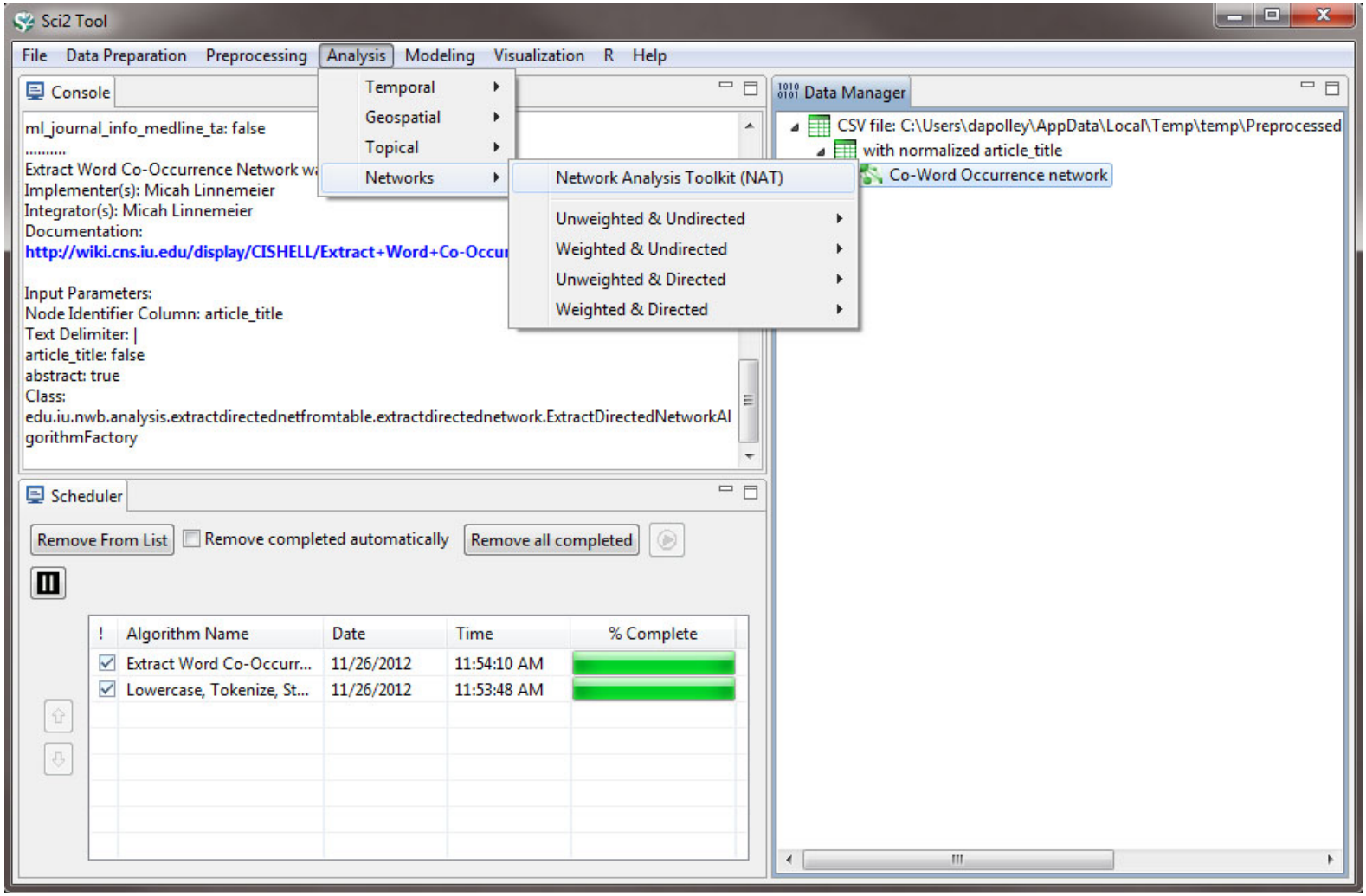
Extracting Word Co-Occurrence Network from SDB Data

Run *Extract Word Co-Occurrence Network* and set the parameters as shown below…



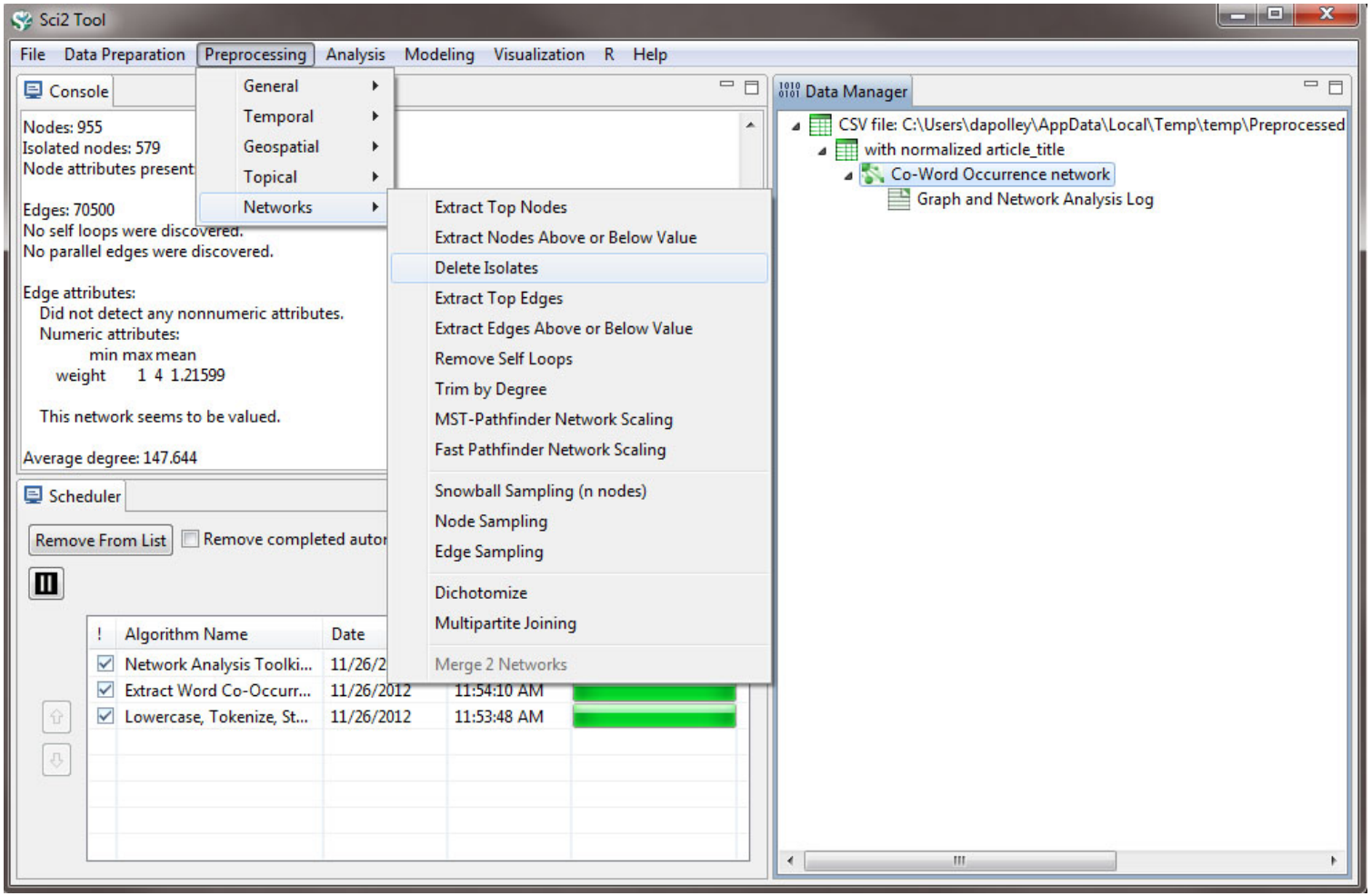
Extracting Word Co-Occurrence Network from SDB Data

To see more information about your network run *Analysis > Networks > Network Analysis Toolkit*...



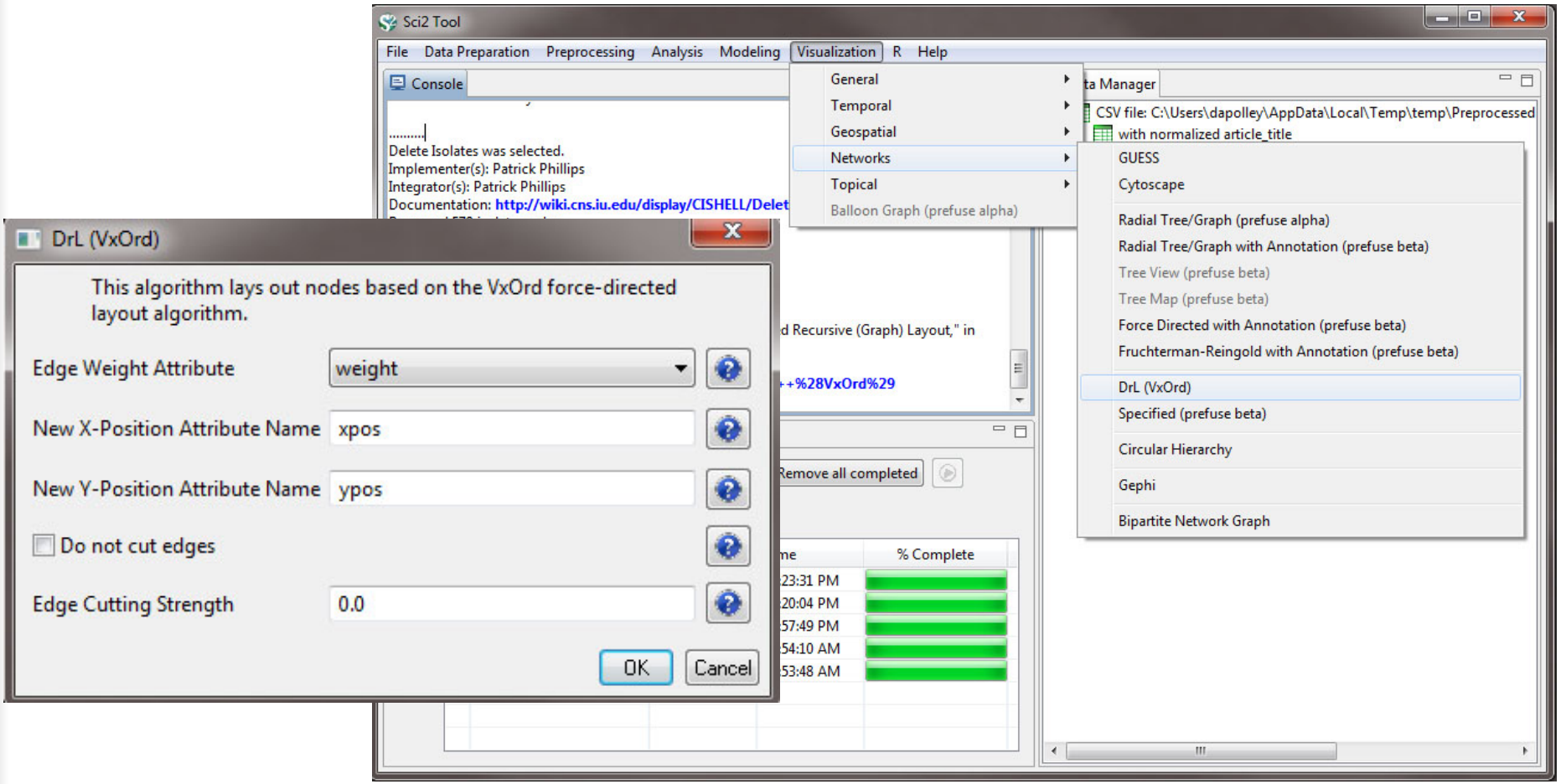
Extracting Word Co-Occurrence Network from SDB Data

Look at the resulting network with the Network Analysis Toolkit. Delete the isolate nodes by running *Preprocessing > Networks > Delete Isolates*



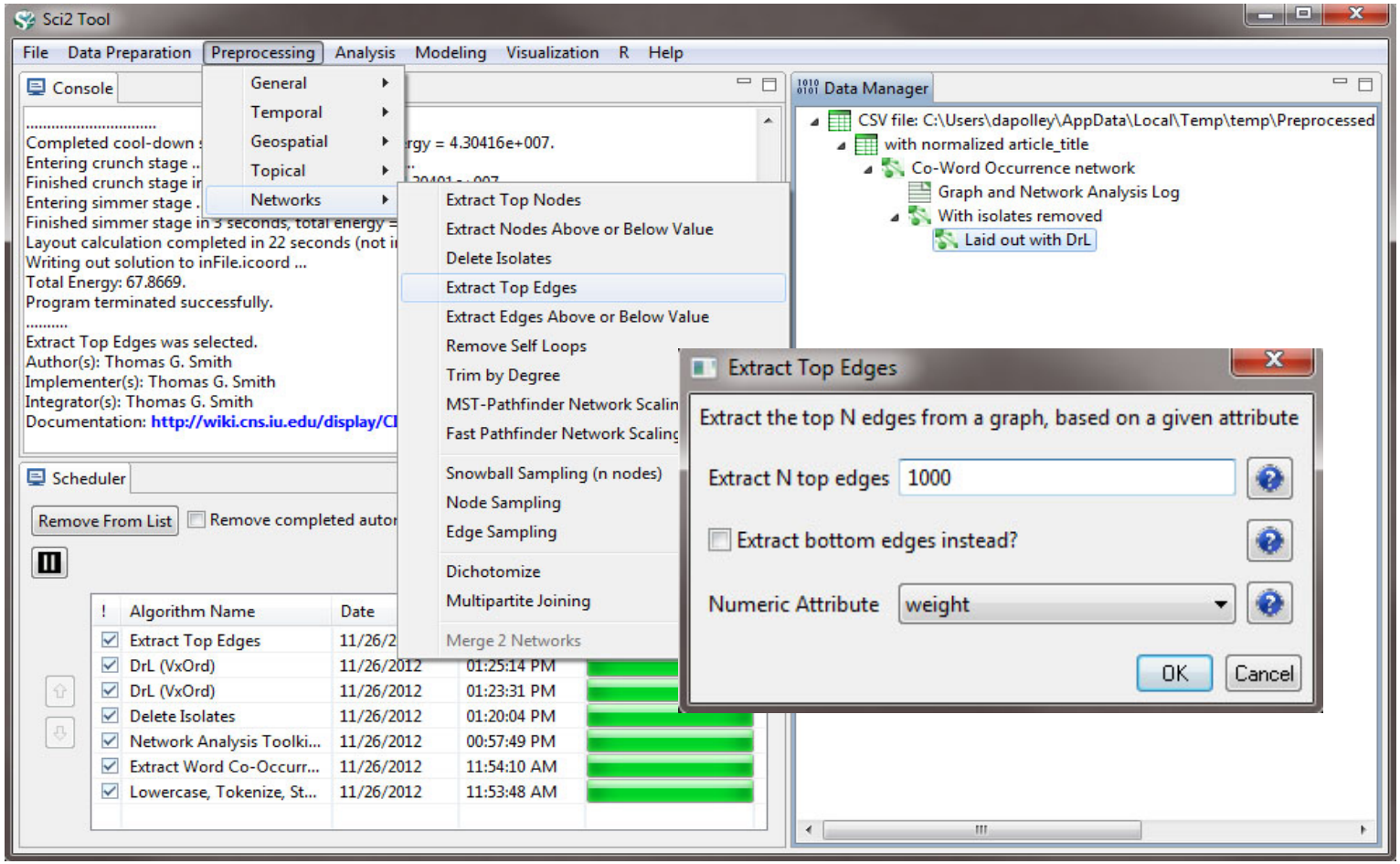
Extracting Word Co-Occurrence Network from SDB Data

Apply *Visualization > Networks > DrL (VxOrd)* and words that are similar will be plotted relatively close to each other. Set the parameters to those shown below…



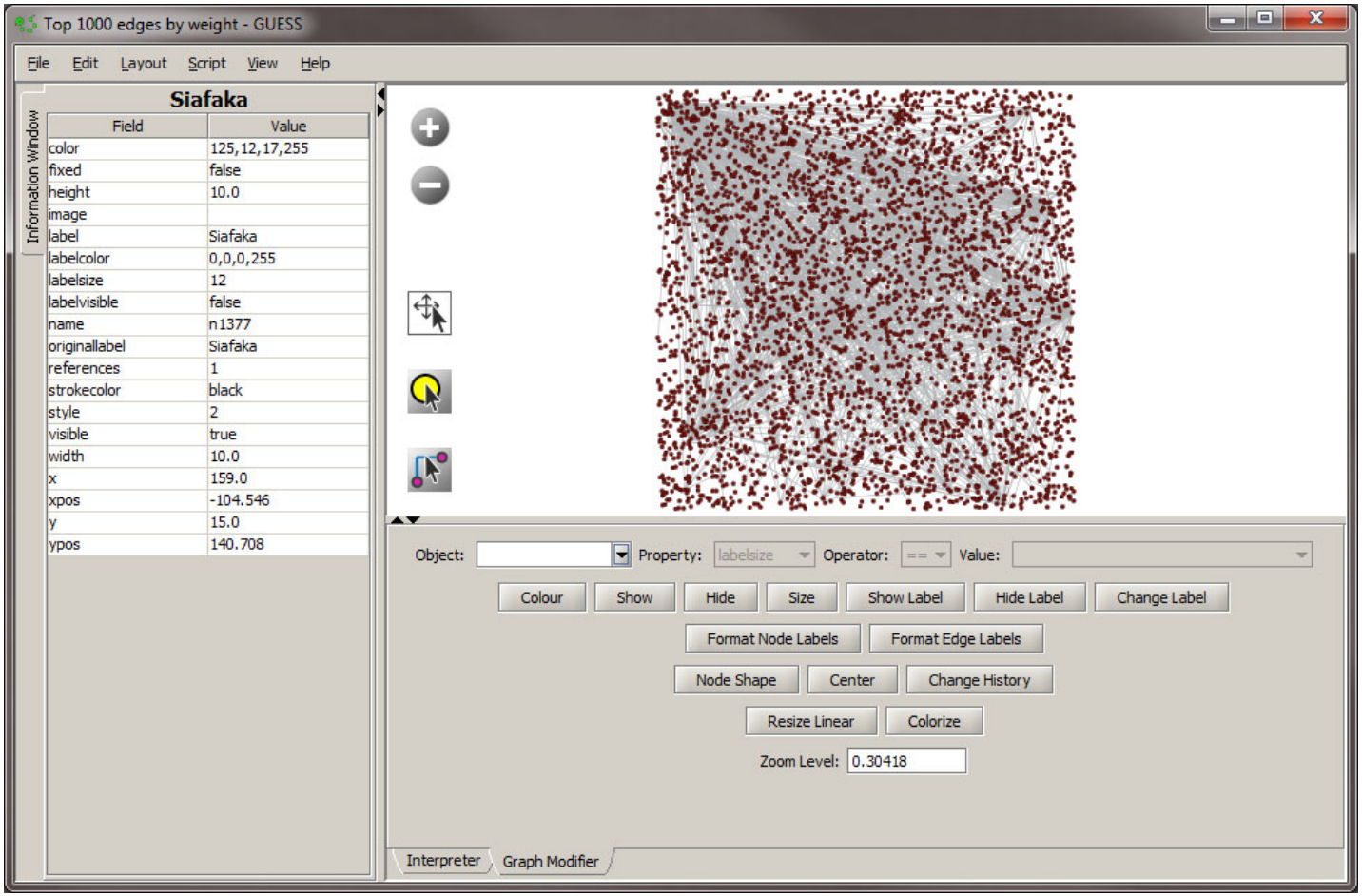
Extracting Word Co-Occurrence Network from SDB Data

Laying out the network with Drl (VxOrd) may take some time, but once the algorithm is complete you will want to keep only the strongest edges, so select the "Laid out with DrL" and run *Preprocessing > Networks > Extract Top Edges* using the parameters shown below…



Extracting Word Co-Occurrence Network from SDB Data

Once edges have been removed, the network "top 1000 edges by weight" can be visualized by running *Visualization > Networks > GUESS*...



Extracting Word Co-Occurrence Network from SDB Data

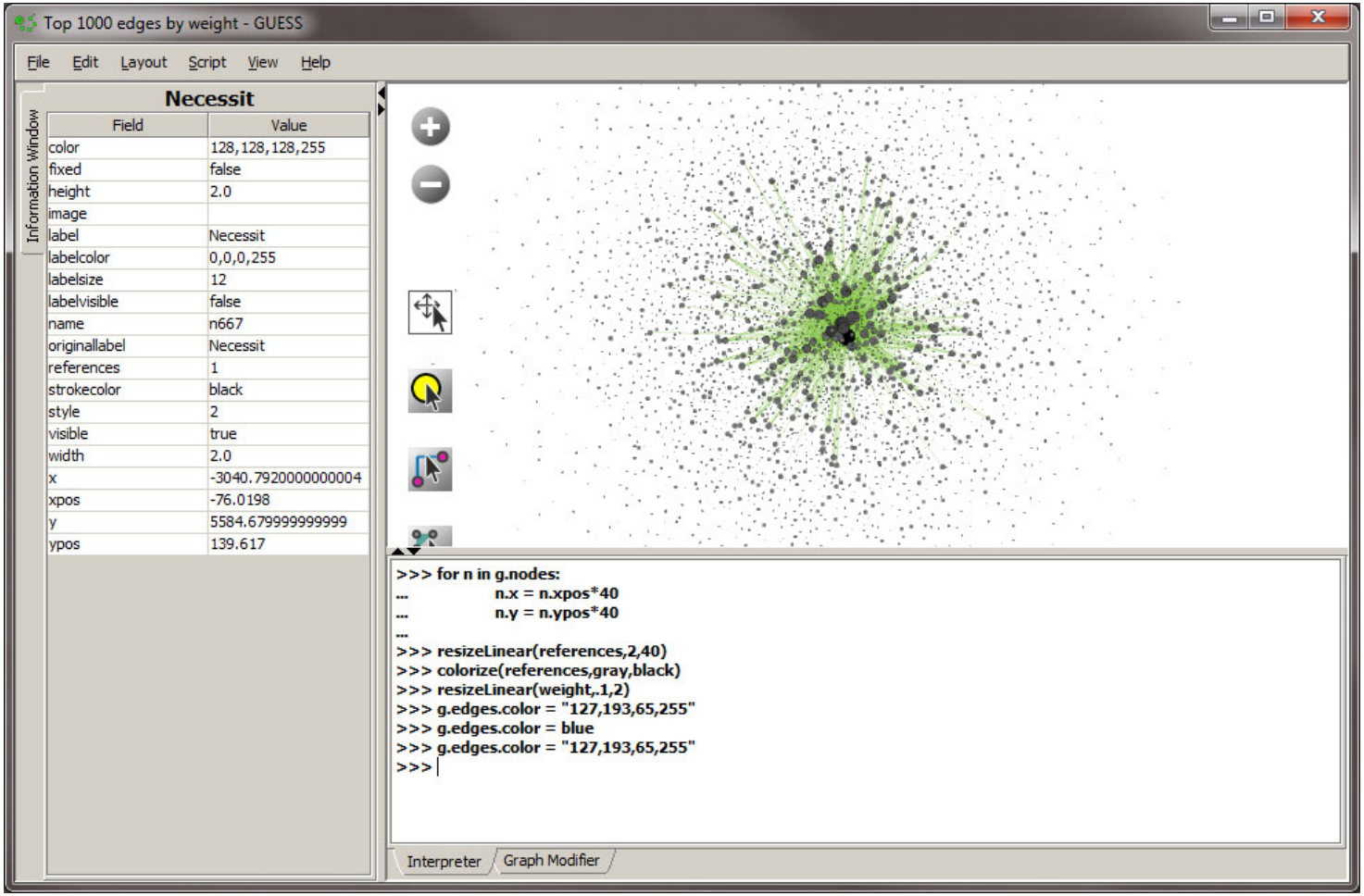
In order to make use of the DrL (VxOrd) force directed layout we applied, we need to change to the interpreter at the bottom of the screen and type in the following commands…

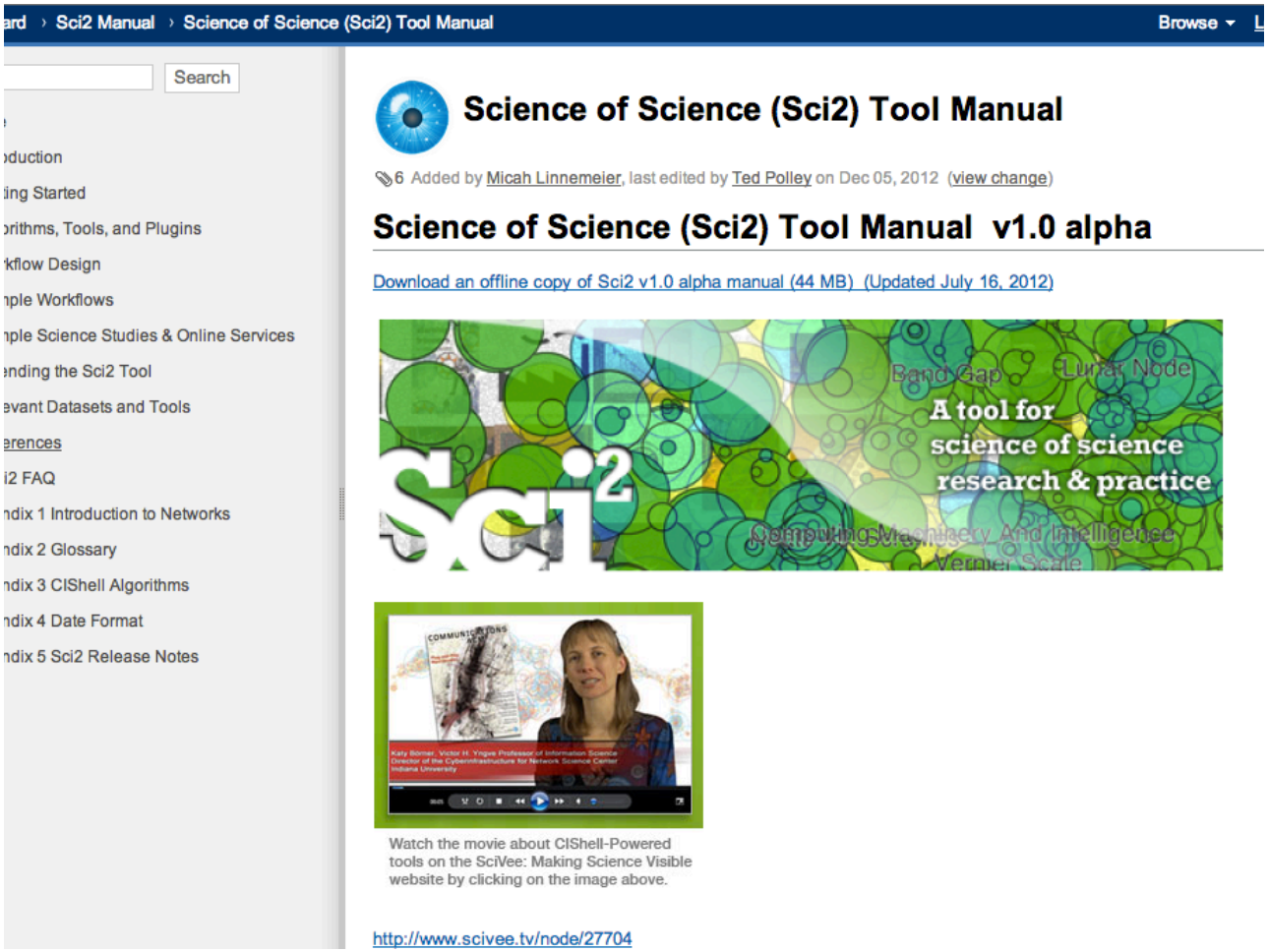
```
>>> for n in g.nodes:
...             n.x = n.xpos*40
...             n.y = n.ypos*40
...
>>> resizeLinear(references,2,40)
>>> colorize(references,gray,black)
>>> resizeLinear(weight,.1,2)
>>> g.edges.color = "127,193,65,255"
>>>
```

Interpreter | Graph Modifier

Extracting Word Co-Occurrence Network from SDB Data

Note, GUESS will not necessarily display the graph in the middle of the screen, you may have to scroll around the screen to find the graph.




ard › Sci2 Manual › Science of Science (Sci2) Tool Manual
Browse
Search
oduction
ing Started
orithms, Tools, and Plugins
kflow Design
mple Workflows
mple Science Studies & Online Services
nding the Sci2 Tool
evant Datasets and Tools
erences
i2 FAQ
ndix 1 Introduction to Networks
ndix 2 Glossary
ndix 3 CIShell Algorithms
ndix 4 Date Format
ndix 5 Sci2 Release Notes
Science of Science (Sci2) Tool Manual
6 Added by Micah Linnemeier, last edited by Ted Polley on Dec 05, 2012 (view change)
Science of Science (Sci2) Tool Manual v1.0 alpha
Download an offline copy of Sci2 v1.0 alpha manual (44 MB) (Updated July 16, 2012)
Sci2
A tool for
science of science
research & practice
Watch the movie about CIShell-Powered tools on the SciVee: Making Science Visible website by clicking on the image above.
http://www.scivee.tv/node/27704

Just the beginning – check the Sci2 website for more!